Improving automatic earthquake detection in the Krafla seismic network

Iñigo Sevilla Echeverría

Faculty of Earth Sciences
University of Iceland
2019

Improving automatic earthquake detection in the Krafla seismic network

Iñigo Sevilla Echeverría

60 ECTS thesis submitted in partial fulfillment of a
Magister Scientiarum degree in Geology

MS Committee
Hanna Blanck
Halldór Geirsson

Master's Examiner
Benedikt Halldorson

Faculty of Earth Sciences
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, October 2019

Improving automatic earthquake detection in the Krafla seismic area
60 ECTS thesis submitted in partial fulfillment of a *Magister Scientiarum* degree in Geology

Faculty of Earth Sciences
School of Engineering and Natural Sciences
University of Iceland
Sæmundargata 2
101, Reykjavik
Iceland

Telephone: 525 4000

Bibliographic information:
Iñigo Sevilla Echeverría, 2019, *Improving automatic earthquake detection for the Krafla seismic area* ,
Master's thesis, Faculty of Earth Sciences, University of Iceland, pp. XX.

Reykjavik, Iceland, October 2019

Abstract

The monitoring of geothermal areas in Iceland is very important, as these are one of the main sources of electricity and hot water. For this purpose, Iceland Geosurvey (ISOR) runs seismic networks for the power companies in most of Icelandic geothermal areas, including Krafla.

The short time average / long time average (STA/LTA) algorithm used for earthquake detection can be adjusted by certain user-set parameters. With the aim of improving the monitoring of the Krafla geothermal area, parameter tuning in the SeisComP3 detection system was made by brute force calculation of thousands of possibilities. This was done by manually checking the traces and writing the picking times for each of the stations for selected time periods, and comparing it to the brute-force calculations carried out using a python script. A set of optimal parameters for the Krafla geothermal area was then selected from the top results of each station, also considering that similar parameters between the stations are advantageous for simplicity reasons.

The results showed that even though the selected set of parameters has been enhanced the detection rate is considerably lower than with ISOR defaults. This can be explained by a limitation of minimum P phases that SeisComP3 has, and as a consequence events that are only detected in less than 6 stations are omitted. The conclusion is that the selected set of parameters proves to be unpractical to implement and under the current circumstances ISOR default parameters are more appropiate. However, significant improvements could be done if SeisComP3 limitation could somehow be bypassed.

Útdráttur

Vöktun jarðhitasvæða á Íslandi er mikilvæg, þar sem umtalsverður hluti rafmagns og hitaveitu kemur frá þessum svæðum. Íslenskar Orkurannsóknir (ÍSOR) sjá um að reka net jarðskjálftamæla fyrir orkufyrirtæki á flestum háhitasvæðum landsins til vöktunar og rannsókna á jarðhitakerfunum.

Algrím sem byggir á að bera saman langtíma- og skammtíma meðaltöl (STA/LTA) jarðskjálftarita eru oft notuð til að bera nema sjálfvirkt hvað er jarðskjálfti innan um annað jarðsuð. Algrímið stjórnast af ákveðnum stikum og hér er skoðað sérstaklega hvernig þessir stikar hafa áhrif á næmni jarðskjálftamælanetsins í Kröflu. Nokkur þúsund mismunandi samsetningar stika voru prófaðar fyrir hverja stöð með hugbúnaði í python. Jarðskjálftarit fyrir ákveðið tímabil voru könnuð handvirkt, komutímar bylgna skráðir og bornir saman við niðurstöður fengnar með mismunandi stikum. Heppilegasta samsetning stika fyrir hverja mælistöð var valin, þar sem forðast var að nema of marga falska atburði, og einnig var leitað að heppilegustu stikum fyrir allt jarðskjálftamælakerfið.

Niðurstöðurnar sýndu að jafnvel besta samsetning stika skilaði sér í verri næmni en upphaflega. Þessa athugun má skýra með takmörkun SeisComP3 að p-bylgja þarf að sjást á að minnsta kosti 6 stöðvum. Því henta þeir stikar sem best komu út ekki vel í núverandi takmörkun SeisComP3, betra væri að leyfa skilgreiningu atburðar með færri p-fösum.

Table of Contents

List of Figures 8

List of Tables 10

Abbreviations 12

Acknowledgements 13

1. Introduction 14

2. Seismology 15
    2.1. Body waves 15
    2.1.1. P waves 15
    2.1.2. S waves 16
    2.2. Surface waves 17
    2.3. Locating earthquake hypocenters using body waves and amplitudes 18

3. Earthquake detection methods 21
    3.1. Short time average / long time average (STA/LTA) 21
    3.2. Waveform cross-correlation / Template matching 23
    3.3. Autocorrelation 24
    3.4. Fingerprint And Similarity Thresholding (FAST) 25

4. Regional area 27
    4.1. Geological setting 27
    4.2. The Krafla geothermal field 28
    4.3. Seismicity in Krafla 30

5. SeisComP 33

6. Methodology 36
    6.1. Python script 36
    6.1.1. Selection of the parameters range 36
    6.1.2. Selection of the signal filter 36
    6.1.3. Reference picks and error 37
    6.1.4. Final output 37
    6.1.5. Selection of optimal parameters 38
    6.1.6. Comparing the results of ISOR default and optimized detection parameters 39

7. Results 40

8. Discussion 46

9. Conclusions 51

References 52

Appendix 1 54

Appendix 2 58

List of Figures

Figure 2.1 - Representation of the P-wave ground propagation in a) spring and b) ground (Tarbuck et al. 2005). 15

Figure 2.2 - Representation of the S-wave ground propagation (Tarbuck et al. 2005). 16

Figure 2.3 - Representation of the Love waves ground propagation (Tarbuck et al. 2005).17

Figure 2.4 - Representation of the Rayleigh waves ground motion (Tarbuck et al. 2005). .17

Figure 2.5 - Distance - time representation used to estimate the distance to the epicenter (Tarbuck et al. 2005). 18

Figure 2.6 - Typical seismogram. A 5 minutes gap between P and S waves can be observed. (Tarbuck et al. 2005). 18

Figure 2.7 - Location of an epicenter based on 3 different stations. (Tarbuck et al. 2005)..19

Figure 3.1 - Illustration of functions and variables of the STA/LTA trigger calculations. a) shows an incoming continuous seismic signal (filtered) as well as the trigger active state between red (trigger on) and blue (trigger off) vertical lines; b) shows the ratio between an averaged absolute signal in the STA and LTA windows, respectively. In this example, the trigger threshold level parameter (red horizontal dotted line) was set to 5 and the detrigger threshold level (blue horizontal dotted line) to 0.5. One can see that the trigger becomes active when the STA/LTA ratio value exceedes 5. It is deactivated when the STA/LTA ratio value falls below 0.5. 22

Figure 3.2 - Autocorrelation schematic technique. A time window (1st column) is autocorrelated with other time windows, mismatching when the waveforms are dissimilar (2nd column), and matching when they are similar (3rd column) (Brown et al. 2008).....24

Figure 4.1 - The KFS and the NVZ. Green frame in inserted figure shows the location of the Northern Volcanic Zone in Iceland (Hjartardottir et al. 2012). 27

Figure 4.2 - General host rock lithology (patterns) and alteration mineralogy (colors) of the geothermal system. Major subvertical faulting denoted by thick lines (Pope et al. 2016). .28

Figure 4.3 - Schematic overview of the Krafla power station (Juliusson et al. 2005). 29

Figure 4.4 - Map of the stations within Krafla and þeistareykir areas (Blanck et al. 2016). ... 32

Figure 5.1 - Processing flow chart of SeisComP3. Screenshots are taken from SC3 Graphical User Interface (GUI), scolv (Behr et al. 2016). 34

Figure 7.1 - Record section from the HVA station. In the upper figure, the filtered waveform is shown. The vertical red lines represent trigger on (not clearly seen due to

overlapping) and vertical blue lines represent trigger off. The lower figure shows the STA/LTA ratio (horizontal lines represent trigger threshold, and horizontal blue lines detrigger threshold).....41

Figure 7.2 - Record section of an earthquake recorded at HVA station. In the upper figure, the filtered waveform is shown (vertical red lines represent trigger on and vertical blue lines trigger off). The figure below shows the STA/LTA ratio (horizontal lines represent trigger threshold and horizontal blue lines detrigger threshold). 42

Figure 7.3 - Record section from the HHK station. In the upper figure, the filtered waveform is shown. The vertical red lines represent trigger on (not clearly seen due to overlapping) and vertical blue lines represent trigger off. The figure below shows the STA/LTA ratio (horizontal lines represent trigger threshold and horizontal blue lines detrigger threshold).....42

Figure 7.4 - Record section of an earthquake recorded at HHK station. In the upper figure, the filtered waveform is shown (vertical red lines represent trigger on, and vertical blue lines trigger off). The figure below shows the STA/LTA ratio (horizontal lines represent trigger threshold and horizontal blue lines detrigger threshold). 43

Figure 7.5 - Record section from the LHN station. In the upper figure the filtered waveform is shown. The figure below shows the STA/LTA ratio. The red box shows a noisy area. examined in figure 7.6 44

Figure 7.6 - Closer look over the noisy area from HHK station. In the upper figure the filtered waveform is shown. The figure below shows the STA/LTA ratio. 44

Figure 8.1 - Example of an event that was omitted using the new parameters and detected by the old ones. The green ticks represent the traces where the event is clear and the red crosses represent those traces where it is not, and could be omitted by SeisComP3.....46

Figure 8.2 - Relationship between the percentage of correct picks and the number of total picks. Each point represents a set of parameters, although different set of parameters can sometimes produce the same results. This data comes from HVA station, where the number of manually selected correct picks was 25. 47

Figure 8.4- Visualization of the picks framework. 50

List of Tables

Table 3.3 - Summary of performance comparison between autocorrelation and FAST for several metrics (Yoon et al. 2015).....26

Table 4.1 - Recorded earthquakes related to injection from November 2014 to October 2015 (Blanck et al. 2016). 30

Table 4.2 - Technical information about the stations (Blanck et al. 2016).....31

Table 6.1 - Reduced sample of the obtained results for station HHK. 38

Table 6.2 - Current set of parameters used at ISOR. 39

Table 7.1 - Final set of parameters obtained for all active stations within the Krafla area, after following the path described in the previous chapter. 40

Table 7.2 - Comparison between results obtained using the new and the ISOR default parameters. 45

Table 8.1 - Example of underpicking set of parameters from station HHK.....48

Table 8.2 - Example of optimal picking set of parameters from station HHK. 48

Table 8.3 - Example of an overpicking picking set of parameters from station HHK. 49

Table 8.5 - Number of events recorded after decreasing and increasing the trigger value. .50

Abbreviations

- MAR - Mid Atlantic Ridge
- SC3 - SeisComP3
- ISOR - Iceland Geosurvey
- STA/LTA - Short time average / long time average
- GUI - Graphical User Interface
- FAST - Fingerprint and Similarity Thresholding
- CC - Correlation coefficient
- LSH - Locality-sensitive hashing
- KFS - Krafla fissure swarm
- NVZ - Northern volcanic zone
- TCP/IP - Transmission Control Protocol / Internet Protocol
- EEW - Earthquake early warning

Acknowledgements

I want to express my gratitude to my supervisors Hanna Blanck and Halldór Geirsson for providing me with the necessary guidance and advice to carry out this research. I also want to thank Benedikt Halldorsson, from Veðurstofa Islands for his feedback and comments about this thesis.

Special mention to Palmi Erlendsson, former employee from Iceland GeoSurvey (ISOR) for his technical advice and help with SeisComP3 at the most struggling moments. He pointed me in the right direction and I gained very valuable knowledge as a result.

Finally, I want to thank my friends and family for their precious support, encouragement and motivation to complete this thesis.

1. Introduction

Iceland is a volcanic island located on top of a mantle plume over the Mid Atlantic Ridge (MAR). Because of its very particular geological and geographical location Iceland has a wide range of geological features, such as glaciers and volcanoes, as well as a number of high temperature geothermal areas. Geothermal areas are very important for the Icelandic society as it provides around 55% of the total consumption of primary energy in Iceland (Loftsdottir et al. 2006) and they are the main source of heating and hot water.

Due to the relevance of the geothermal areas, it's very important to develop a safe and sustainable way to exploit them. Iceland Geosurvey (ISOR) is an Icelandic company established in 2003, which focuses on geothermal exploration, development and utilization, although they cover other geoscience related fields as well (ISOR 2019). ISOR, in cooperation with different companies, is operating over several geothermal fields in Iceland such as Reykjanes, Svartsengi, Krafla, Námafjall - Bjarnarflag and Theistareykir. Monitoring of all geothermal fields is recommended, and for that purpose ISOR is running local seismic networks.

The software used by ISOR to record and process data from their seismic networks is SeisComP3 (SC3), which applies a short time average / long time average (STA/LTA) algorithm for earthquake detection. The STA/LTA algorithm has different variables that can be adjusted for enhanced results.

This research focuses on the Krafla geothermal area and its seismic network run by ISOR. The problem is that with the current STA/LTA settings used at ISOR could be improved to reduce false triggering. Human re-evaluation of the detected earthquakes helps to discriminate between the real and the false events, but can be very time consuming when the number of detected earthquakes is high.

The goal with this thesis is to improve the automatic earthquake detection at the Krafla seismic network, modifying the variables from the STA/LTA algorithm used by SC3. The methodology selected is a brute-force calculation of thousands of different combinations of STA/LTA variables for each of the stations within the Krafla seismic network.

2. Seismology

In this chapter we will review seismological concepts, with the aim of providing the reader with the essential knowledge needed to understand this research. An earthquake is a release of energy that occurs by the sudden breaking of rock. Earthquakes can also be triggered or induced by human activity, for example by explosions. The energy propagates through the earth in the form of seismic waves and is recorded on seismograms.

There are two main different types of seismic waves. The first type are body waves (travel through the earth inner layers) and the second are surface waves (can only move along the surface of the planet).

2.1. Body waves

Traveling through the inner part of the earth, body waves arrive before surface waves emitted by an earthquake. Body waves are of a higher frequency than surface waves. There are two different types of body waves: P waves and S waves.

2.1.1. P waves

The first type of body wave is the primary wave or P wave. It is the fastest seismic wave and thus the first one to arrive at a seismic station. The P wave can propagate through solid rocks and fluids (water or other liquids layers inside the earth). P waves are also known as compressional waves, because of their horizontal push and pull movement (represented in figure 2.1). Subjected to a P wave, particles move in the same direction as the wave is propagating. They have a smaller amplitude than the S waves and are unlikely to be felt by humans.

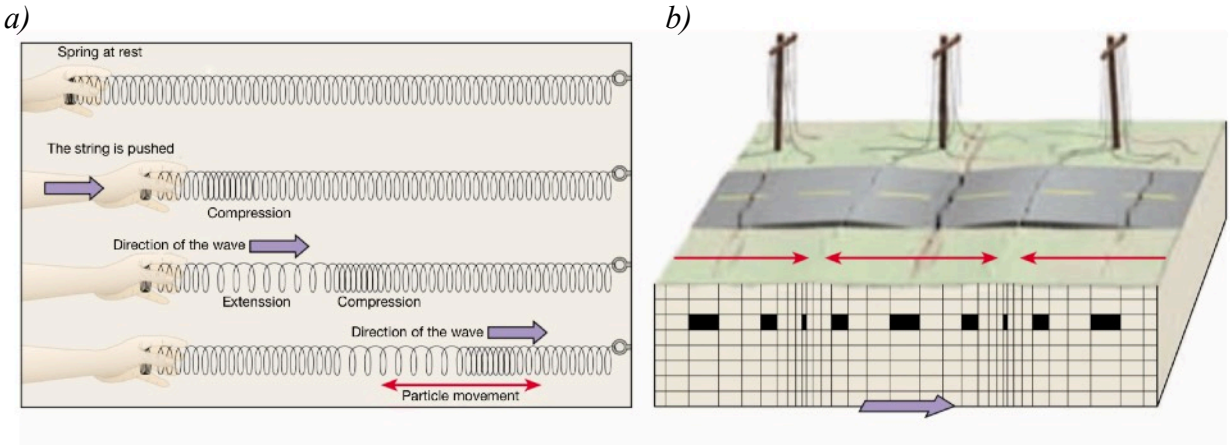

Figure 2.1 - Representation of the P-wave ground propagation in a) spring and b) ground (Tarbuck et al. 2005).

The velocity of the P waves (v_p) is given by equation 2.1:

$$v_p = \sqrt{\frac{K + \frac{4}{3} \cdot \mu}{\rho}}$$ Equation 2.1

,where K is the bulk modulus, μ is the shear modulus and ρ is the density of the material the wave is traveling through.

2.1.2. S waves

The second type of body wave is the shear waves or S waves, which is the second wave that is recorded by the seismometer. S waves are slower than P waves and can only move through solid material, not through liquids as their shear modulus is zero. The movement of S waves (figure 2.2) is perpendicular to the direction that the wave is travelling (up and down or side to side)

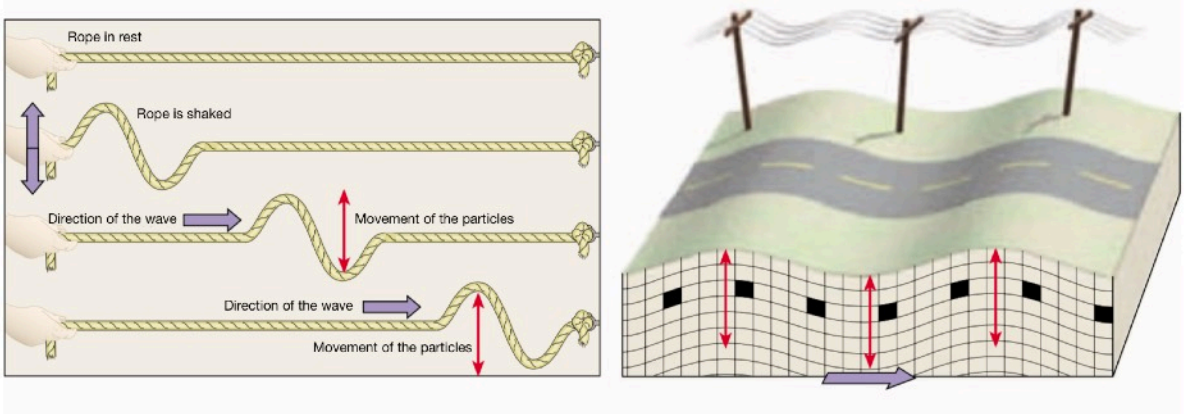

Figure 2.2 - Representation of the S-wave ground propagation (Tarbuck et al. 2005).

The velocity of the S waves is given by the following formula:

$$v_s = \sqrt{\frac{\mu}{\rho}}$$ Equation 2.2

,where μ is the shear modulus of the elastic material and ρ is the density.

2.2. Surface waves

Surface waves travel through the crust. They have lower frequency than body waves, and sometimes they can be recognized in seismograms. They arrive later than body waves but they cause almost all the damage and destruction associated with earthquakes. They are specially relevant in shallow and strong earthquakes due to their bigger amplitude and their importance is reduced at deep and weaker earthquakes. There are two different types of surface waves: Love waves and Rayleigh waves.

Love waves are the fastest surface waves and are confined to the surface of the crust, producing entirely horizontal motion (Figure 2.3).

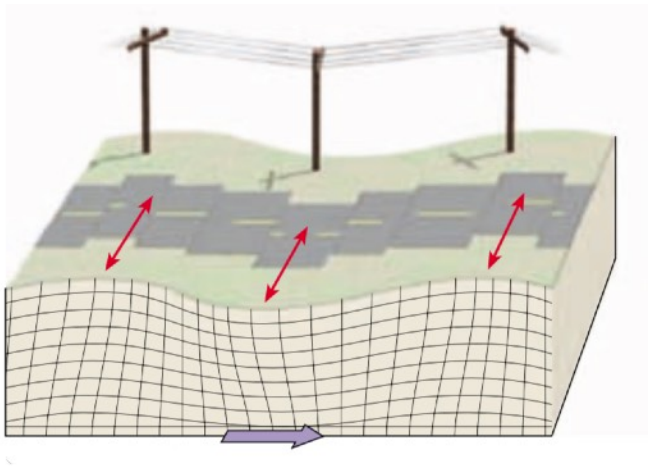

Figure 2.3 - Representation of the Love waves ground propagation (Tarbuck et al. 2005).

Rayleigh waves roll across the ground just like a wave rolls across the ocean, moving the ground up and down and side to side in the same direction as the wave is propagating (figure 2.4). Most of the shaking felt from an earthquake is due to the Rayleigh waves, which can be of higher amplitude than the other wave types.

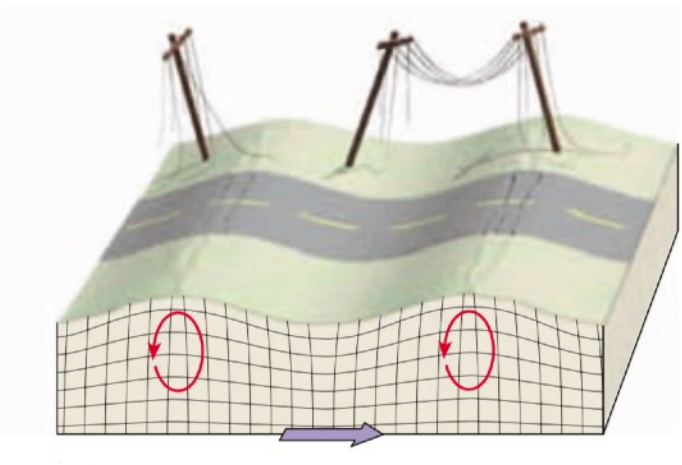

Figure 2.4 - Representation of the Rayleigh waves ground motion (Tarbuck et al. 2005).

2.3. Locating earthquake hypocenters using body waves and amplitudes

The speed difference between P and S waves gives us a method to locate the epicenter. P waves are faster than S waves, and will always arrive earlier and bigger time gaps indicate bigger distance to the earthquake. By using earthquakes of which epicenter is well known (for example nuclear blasts), we can derive a time-distance relationship (figure 2.5).

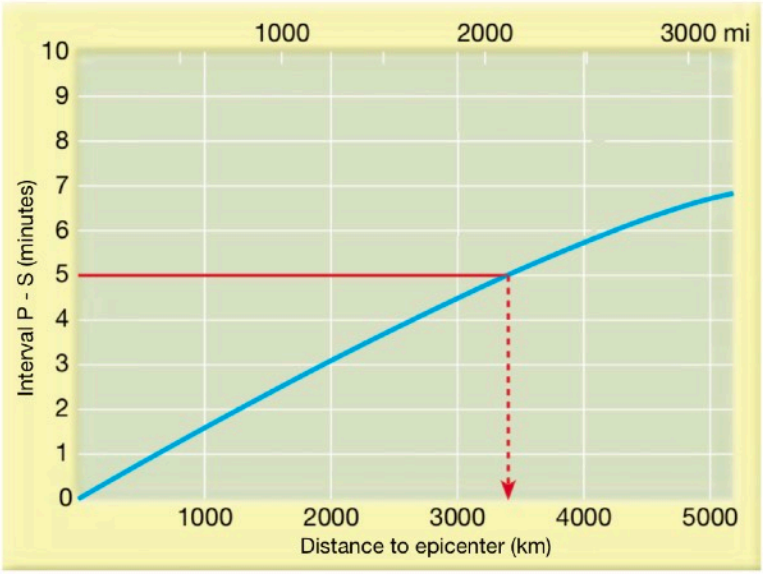

Figure 2.5 - Distance - time representation used to estimate the distance to the epicenter (Tarbuck et al. 2005).

Let's consider for example the seismogram shown in figure 2.6. There is a time gap of approximately 5 minutes between the arrival of the P waves and the S waves, considering a pick as the moment of arrival of P and S waves respectively. So, according to the distance-time relationship, we can estimate that the epicenter is approximately 3.400 km away. (Tarbuck et al. 2005).

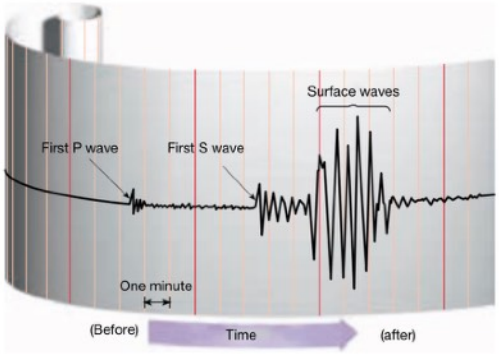

Figure 2.6 - Typical seismogram. A 5 minutes gap between P and S waves can be observed. (Tarbuck et al. 2005).

After we know the distance, we also need to know the direction of the earthquake. This can be done with a simple triangulation method, by measuring the distance in 3 or more stations as it is shown on figure 2.7.

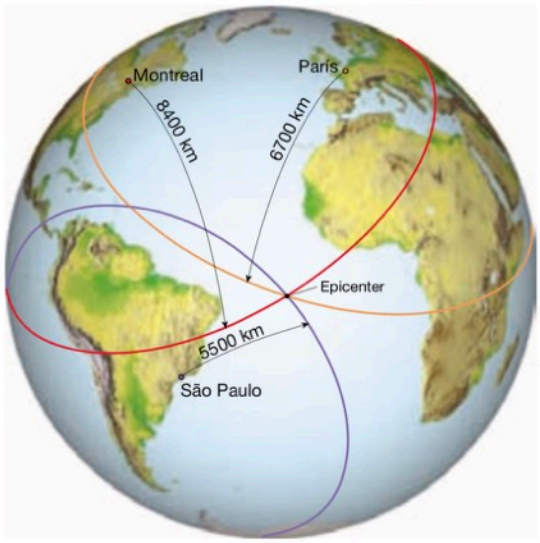

Figure 2.7 - Location of an epicenter based on 3 different stations. (Tarbuck et al. 2005).

The method for earthquake location explained above is a very direct and straightforward way to do it. However, different ways also exists. The software used for this research (SC3) uses a different approach, considering P arrival times and their associated envelopes and amplitudes (SC3 documentation). The envelope of a seismic signal is a smooth curve outlining its extremes. Thus, it gives an approximation of a constant amplitude.

The amplitude of the seismic waves is higher near the epicenter. As the waves propagate they attenuate until they cannot be felt anymore. The amplitude of the seismic waves could be used for earthquake location in similar ways as S waves. A short time difference between P and S waves indicate a close epicenter. Similarly, an event with a higher amplitude of the seismic waves at a certain station could indicate a closer epicenter than in a different station with lower amplitudes. For this method to be accurate, we need to make sure that picks are correctly grouped together before declaring and locating an event. If many P picks and associated amplitudes from different stations are detected within a certain small period an event will be triggered.

For local earthquakes, regional geology plays a more important role, as it may affect the local velocity structure and travel times of the seismic waves more pronouncedly.

In SC3, several modules work together to produce arrival times or picks (scautopick) and their associated envelopes and magnitudes (scenvelope and scmag). Scautoloc groups picks to locate events. Further information about SC3 and their modules can be found in chapter 5.

3. Earthquake detection methods

Automatic detection of earthquakes in continuous streams of data from a network of seismic stations is challenging because earthquake signals are often embedded within seismic noise. Amplitude threshold trigger algorithms measure the amplitude of the recorded seismic signal and trigger when it exceeds a certain limit. Although they can be useful for detecting the strongest earthquakes, they are not effective with events camouflaged in seismic noise, and more advanced methods are required. Short time average / Long time average (STA/LTA) is the most popular one, but newer approaches like Fingerprint and Similarity Thresholding (FAST) are appearing. In this chapter, different options such as STA/LTA, cross-correlation, autocorrelation and FAST are explained.

3.1. Short time average / long time average (STA/LTA)

The short time average / long time average (STA/LTA) trigger is commonly used in weak motion applications that try to record as many seismic events as possible. These are the applications where the STA/LTA algorithm is most useful. It can also be used for strong motion applications, except when the interest is only the strongest earthquakes. The STA/LTA algorithm detects earthquakes by comparing short term levels of ground shaking to longer term levels of ground shaking, significantly improving the recording of weak earthquakes in comparison to amplitude threshold trigger algorithms. It also decreases the number of false records triggered by natural and man-made seismic noise.

The STA/LTA algorithm continuously keeps track of the always-present changes in the seismic noise amplitude at the station and automatically adjusts the seismic station's sensitivity to the actual seismic noise level. As a result, a significantly higher sensitivity of the system during seismically quiet periods is achieved and an excessive number of falsely triggered records is prevented, or at least mitigated, during seismically noisy periods.

The STA/LTA requires four different variables, a short time average window (STA), a long time average window (LTA), a trigger threshold and a detrigger threshold. All variables are explained below and illustrated in figure 3.1:

- *STA*: the short time average measures the ‘instant’ value of a seismic signal (its envelope). Normally, the STA window should be longer than a few periods of a typically expected seismic signal. If the STA window is too short, the averaging of the seismic signal will not be accurate, as it becomes influenced by individual periods of the seismic signal. On the other hand, STA window length must be shorter than the shortest event duration we want to capture.

The STA acts as a signal filter. The shorter the selected duration, the higher the trigger sensitivity to short lasting local earthquakes compared to long lasting and lower frequency distant earthquakes. Therefore, by changing the STA window length, we can prioritize capturing distant or local events.

The STA window length is also relevant with respect to false triggers. When decreasing the STA window, the trigger gets more sensitive to spike-type man-made noise. At highly noisy sites, one will be frequently forced to make the STA window longer than these spikes if false triggers are numerous. The drawback is that this reduces the sensitivity of the recording and some events may not be detected (Trnkoczy 2012).

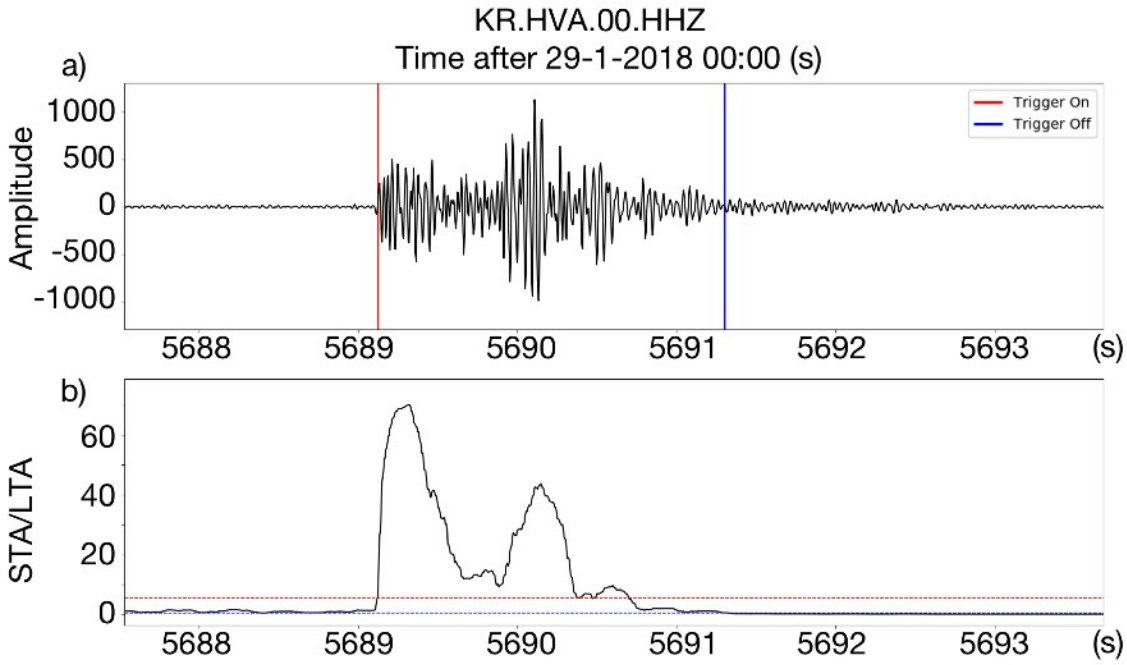

Figure 3.1 - Illustration of functions and variables of the STA/LTA trigger calculations. a) shows an incoming continuous seismic signal (filtered) as well as the trigger active state between red (trigger on) and blue (trigger off) vertical lines; b) shows the ratio between an averaged absolute signal in the STA and LTA windows, respectively. In this example, the trigger threshold level parameter (red horizontal dotted line) was set to 5 and the detrigger threshold level (blue horizontal dotted line) to 0.5. One can see that the trigger becomes active when the STA/LTA ratio value exceedes 5. It is deactivated when the STA/LTA ratio value falls below 0.5.

- *LTA*: the long time average measures average amplitude seismic noise. The window should be longer than a few periods of typically irregular seismic noise fluctuations. A short LTA window length allows the LTA value to adapt to the slowly increasing amplitude of the seismic waves. So, the STA/LTA ration remains small in spite of increasing STA. As a consequence, a short LTA window diminishes trigger sensitivity to distant events with high epicentral distance. On the other hand, using a long LTA window increases the trigger sensitivity to the emergent earthquakes as the LTA value is not so rapidly influenced by the emergent signal (Trnkoczy 2012).

- *Trigger threshold:* The trigger threshold determines which events will be recorded and which not, as it sets the limit STA/LTA ratio over which triggering occurs. The higher value we set, the more earthquakes will be missed, but the fewer false events will result. The lower the STA/LTA trigger threshold value is set, the more sensitive the seismic station will be and the more events will be recorded (Trnkoczy 2012).

- *Detrigger threshold:* When the STA/LTA ratio reaches the trigger threshold, the system is prevented from triggering again until the STA/LTA ratio drops below the detrigger threshold, setting the level under which the triggering routine is re-established. When the detrigger value is too high, triggering routine resets very often and as a result a single event could produce several triggers. On the other hand, if the detrigger value is too low, trigger can be active for a long time, preventing the system from triggering again.

The STA/LTA algorithm calculates first the absolute amplitude of each data sample of an incoming signal, based on its envelope. Next, the average of absolute amplitudes in both windows is computed and a ratio of both values (STA/LTA ratio) is obtained. This ratio is compared to a user selected threshold value – the STA/LTA trigger threshold level. If the ratio exceeds this threshold, a trigger is declared. After the seismic signal gradually terminates, the detrigger occurs. This happens when the current STA/LTA ratio falls below another user selected parameter (STA/LTA detrigger threshold level). The STA/LTA detrigger threshold level should be lower (or rarely equal) than the STA/LTA trigger threshold level.

The STA/LTA trigger parameter settings are always a tradeoff among several seismological and instrumental considerations. The goal when choosing a set of parameters is the highest possible seismic station sensitivity for a given type of seismic signal, could also include the target 'all earthquakes', at a still tolerable number of false triggers.

Different variations of the STA/LTA algorithm have been developed (Withers et al. 1998), such as the recursive STA/LTA and delayed STA/LTA. In some cases, an STA/LTA algorithm is combined with further processing, like separating P and S phases (Wang et al. 2017), improving the results.

3.2. Waveform cross-correlation / Template matching

Waveform cross-correlation (also known as template matching) has proven to be a sensitive, discriminative method for finding a known seismic signal in noisy data. Seismic sources that repeat in time over the course of weeks, months or years have very similar waveforms when recorded at the same station (Schaff et al. 2004). Path effects are almost the same, as studies on time-dependent travel time variations before and after large earthquakes have shown (Schaff et al. 2004). Waveform cross-correlation takes advantage of the resulting waveform similarity to perform as a sensitive earthquake detector.

Waveform cross-correlation has a "one-to-many" search method that computes the normalized correlation coefficient (CC) of a template waveform with successive candidate

time windows of continuous waveform data, and any candidate window with CC values exceeding certain limits is considered a detection. It allows detection of signals with extremely low signal-to-noise ratio, with few false positives when the template includes waveforms from multiple channels and stations (Yoon et al. 2015).

A major disadvantage of waveform cross-correlation is that it requires a previous waveform template, so it has low general applicability. Templates are often chosen by extracting waveforms from catalog earthquakes or by impulsive event waveforms picked from continuous data under human inspection. Waveform cross-correlation is not an effective way to find unknown sources with low signal-to-noise ratio.

3.3. Autocorrelation

Autocorrelation is an exhaustive "many-to-many" search for similar waveforms when the desired signal waveform is unknown. Knowing that the seismic signal of interest has a short duration, the continuous data is partitioned into a number of short overlapping windows and all possible pairs of windows are correlated. Window pairs with correlation coefficients exceeding a detection threshold are marked as candidate events, as shown in figure 3.2. These candidate events can be post-processed with additional cross-correlation or grouped into "families" and stacked to form less noisy template waveforms. Autocorrelation has successfully found both known and previously unknown low frequency earthquakes within tectonic tremor (Brown et al. 2008).

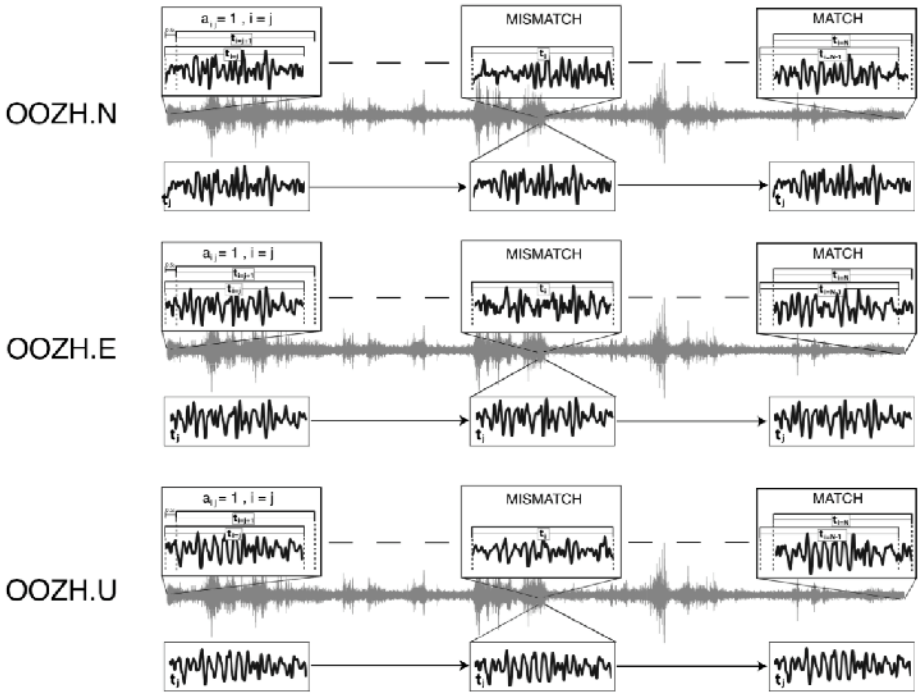

Figure 3.2 - Autocorrelation schematic technique. A time window (1st column) is autocorrelated with other time windows, mismatching when the waveforms are dissimilar (2nd column), and matching when they are similar (3rd column) (Brown et al. 2008).

Autocorrelation has a major disadvantage because it is computationally intensive, making it unsuitable for detecting earthquakes in massive continuous data sets. Autocorrelation

performs a significant amount of redundant work because most pairs of windows are uncorrelated and not of interest for detection and highly similar earthquakes detected by autocorrelation are a tiny fraction of the total number of pairs. It is, however, suitable for detecting frequently repeating earthquakes in relatively small continuous datasets (Brown et al. 2008) where the number of window pairs is small. But the run time of autocorrelation makes it impractical to find infrequently repeating event in days, weeks, months or even years of continuous seismic data over a network of hundred of channels and stations without using large-scale computing resources.

3.4. Fingerprint And Similarity Thresholding (FAST)

A relatively recent approach for earthquake detection was developed by Yoon et al. (2015). Many algorithms have been developed to efficiently search for items in massive data sets, applications include identifying similar files in a large file system, finding near-duplicate Web pages, detecting plagiarism in documents and recognizing similar audio clips for music identification among others. It is possible to develop a fast, efficient, automated blind detection of similar earthquake waveforms in huge volumes of continuous data by leveraging scalable algorithms commonly used in the computer science community.

A widely used method for high-dimensional approximate nearest-neighbors search is “locality-sensitive hashing” (LSH). It allows to avoid comparing dissimilar pairs which constitutes most pairs of waveforms in the data. Instead, LSH returns a shorter list of “candidate pairs” that are likely to be similar. Each item is inserted into one hash bucket that is selected based on the output of a hash function. A hash table contains many hash buckets and the hash function determines how items are distributed among the different hash buckets. So, only pairs of similar items (seismic signals) within the same hash bucket become candidate pairs while pairs of items that do not appear together in the same hash bucket (which comprise most pairs) can be ignored. Therefore, LSH allows search for similar items with a runtime that scales near-linearly with the number of windows from continuous data which is much better than autocorrelation.

Instead of directly comparing waveforms, they first perform feature extraction to condense each waveform into a compact “fingerprint” that retains only its key discriminative features. A fingerprint serves as a proxy for a waveform, thus, two similar waveforms should have similar fingerprints and two dissimilar waveforms should have dissimilar fingerprints. Fingerprints are assigned to LSH hash buckets (instead of waveforms).

Yoon et al. (2015) tested the detection capability of FAST and compared it with autocorrelation (table 3.3). It showed a comparable performance to that of autocorrelation with some additional false positives. However, FAST has an enormous advantage over autocorrelation in terms of runtime. This advantage can be expected to increase for longer duration continuous data sets.

Table 3.3 - Summary of performance comparison between autocorrelation and FAST for several metrics (Yoon et al. 2015)

Metric	Autocorrelation	FAST
1. Total number of detected events	86	89
2. Number of false detections (false positives)	0	12
3. Number and percentage of catalog detections	24/24 = 100%	21/24 = 87.5%
4. Number of new detections from both algorithms	43	43
5. Number of new detections from one, missed by the other	19	25
6. Number of missed detections (false negatives)	25	22
7. Runtime	9 days 13 hours	1 hour 36 min

4. Regional area

4.1. Geological setting

Iceland is located on an insular shelf that is located in the Mid-Atlantic Ridge. Vigorous vulcanism in Iceland caused by a mantle plume has build up a plateau around 3 km higher than the average Mid-ocean ridge height, with 3 to 4 times thicker crust than average oceanic crust (Bjarnason 2008). The rift zones in Iceland are characterized by active volcanism and faults, and mark the present plate boundary between the North American and Eurasian plates. The Krafla Fissure Swarm (KFS) and Krafla central volcano constitute the Krafla volcanic system. It is one of the five volcanic systems (Þeistareykir, Krafla, Fremrinámar, Askja and Kverkfjöll, from north to south) within the Northern Volcanic Zone (NVZ) in Iceland (Figure 4.1).

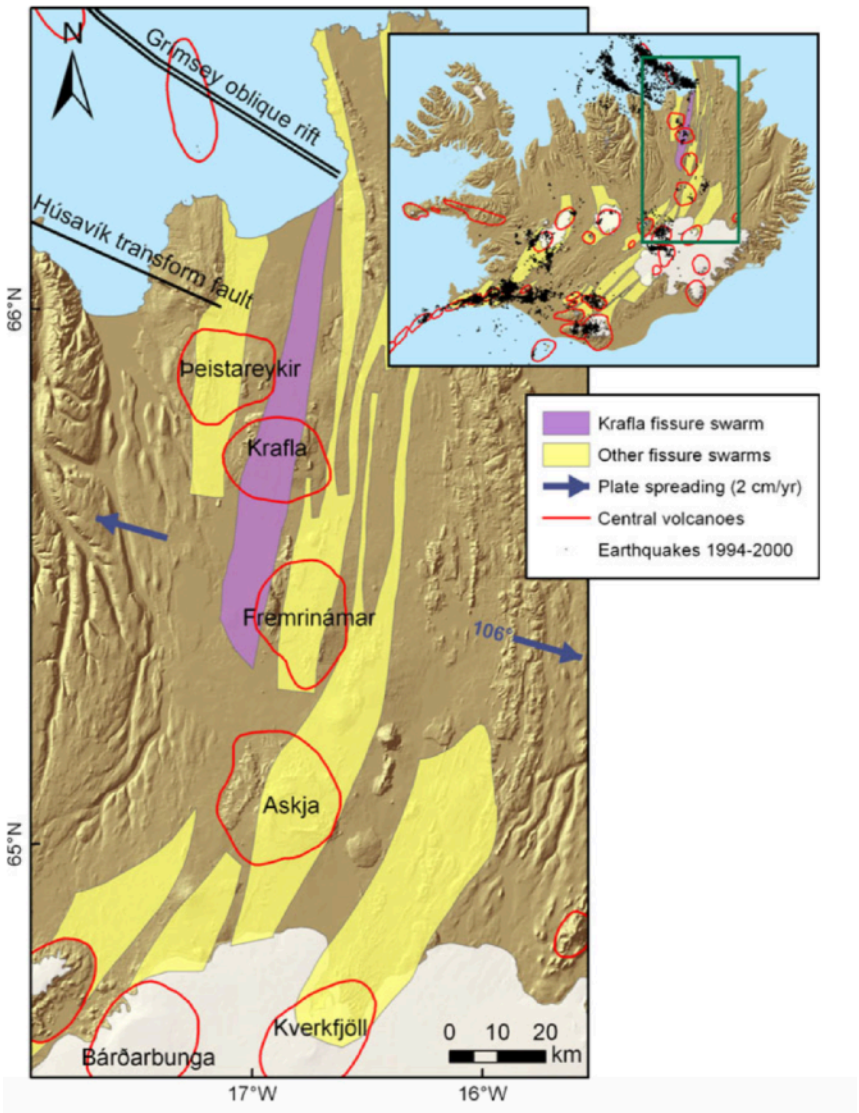

Figure 4.1 - The KFS and the NVZ. Green frame in inserted figure shows the location of the Northern Volcanic Zone in Iceland (Hjartardottir et al. 2012).

Each volcanic system comprises NNE trending fissure swarms that transect their central volcanoes approximately perpendicular to the spreading direction. At the NVZ, the plate boundary diverges at an estimated rate of 2 cm/year (Demets et al. 1994).

The KFS extends around 50 km towards the north and about 40 km towards the south from the Krafla central volcano. The fissure swarm is mostly situated in the post-glacial lava flows emplaced during the last 10.000 years. Fractures within the swarm are mainly oriented N to NNE, with widths up to 40 m. During historical times, two rifting episodes have occurred on the KFS: the 1724-1729 “Myvatn rifting episode” and the instrumentally recorded 1975-1984 “Krafla rifting episode”. Periods of intense seismic activity occurred within the fissure swarm during both episodes. Fissure eruptions accompanied the rifting (Hjartardottir et al. 2012).

The Krafla central volcano extends over an area of 21 by 17 km and exhibits relatively low relief (300-500 m elevation). Since the last glacial period 35 eruptions have been identified within the Krafla volcanic system using tephra deposits, including the Myvatn rifting episode and the Krafla rifting episode. The Krafla rifting episode caused about 8 m horizontal extension in the central part of the caldera (Schuler et al. 2015).

4.2. The Krafla geothermal field

High-temperature geothermal fields are often associated with central volcanoes, where faults and fissures transecting the volcanoes allow water to easily penetrate the shallow hot crust. The Krafla geothermal field is located inside the Krafla caldera. Exploration drilling started in 1974 and the power plant started operation in 1977. A generalized NW-SE geologic cross section is given in figure 4.2. Most wells reach a temperature of 240C at 1500 m depth below sea level. In spring 2009 the exploration well IDDP-1 penetrated rhyolitic magma at 2104 m depth (1551 m below sea level) (Schuler et al. 2015, Pope et al. 2016).

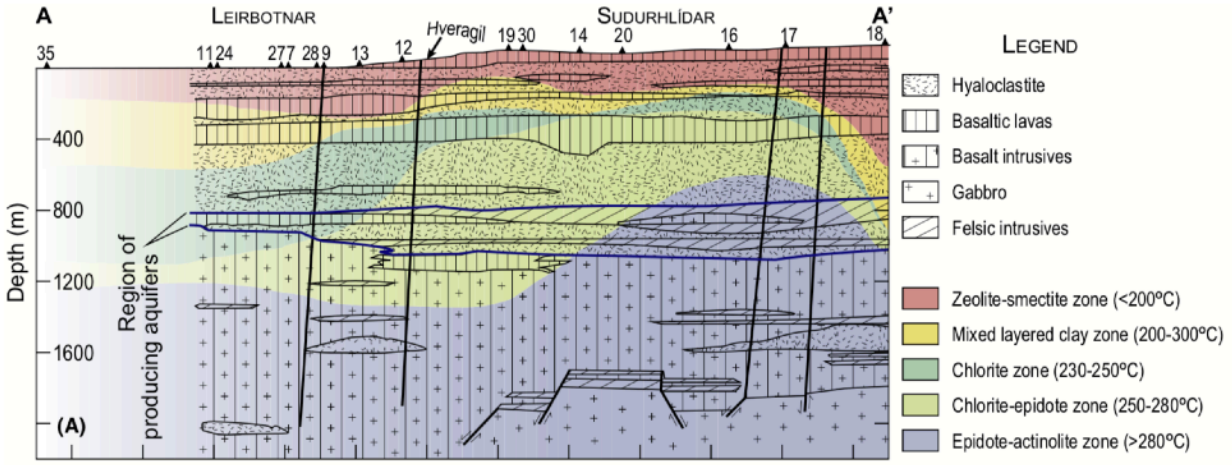

Figure 4.2 - General host rock lithology (patterns) and alteration mineralogy (colors) of the geothermal system. Major subvertical faulting denoted by thick lines (Pope et al. 2016).

Krafla geothermal field is mainly known for the volcanic eruptions during the construction time and the first years of production. A decision to develop the field was made in 1974 by two state owned organizations, Kröflunefnd (power station) and the National Energy Authority (steam supply system), on the basis of exploration wells. Construction of a 60 MW power station started in summer of 1975 but only a few months later, a volcanic eruption series started in the north part of the central volcano. When the eruptions series ended in 1984, magma had been released from the magma chamber 21 times, 9 times resulting in eruptions. Luckily, lava did not flow into the Hlidardalur valley, where the power station is located and the construction continued. Apart from surface elevation changes, the eruptions caused significant changes in fluid chemistry, mostly an increase in corrosive volcanic gasses in the wells located in the Viti area. As a result, wells had to be drilled in less favorable sites, further away from the eruption zone. The developers soon realised that the steam was insufficient for the planned power plant and only one of the turbines was fully installed. A schematic overview of the power station is represented in figure 4.3 (Juliusson et al. 2005).

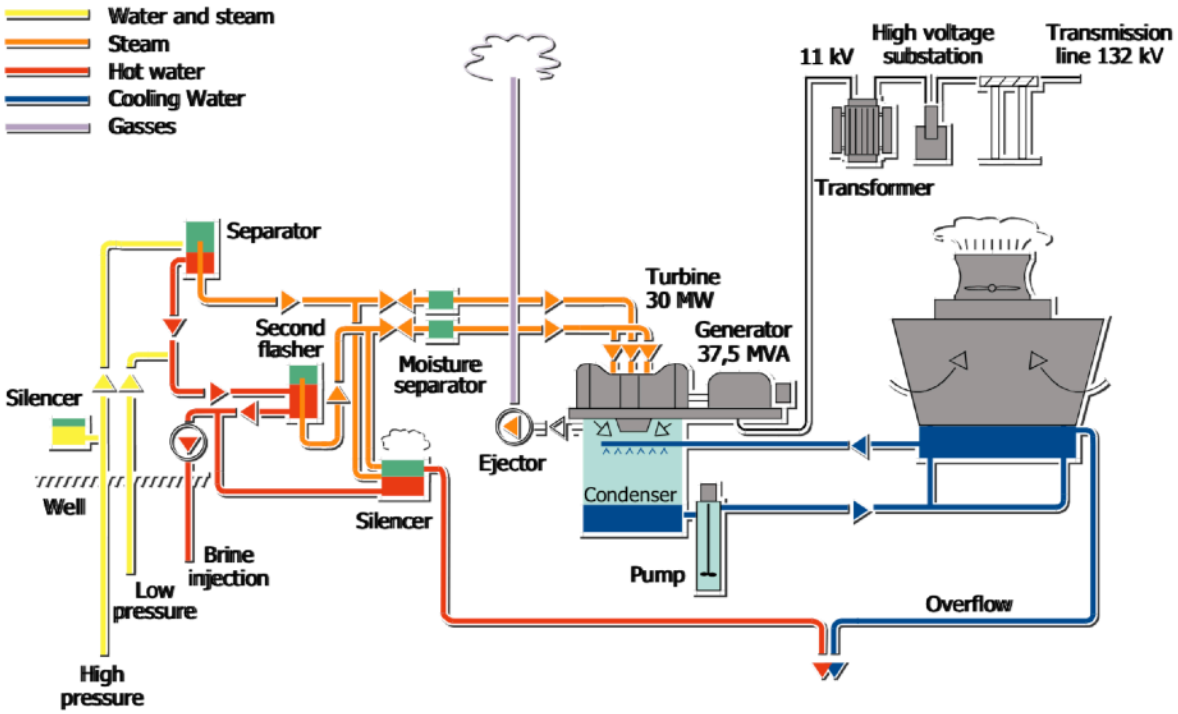

Figure 4.3 - Schematic overview of the Krafla power station (Juliusson et al. 2005).

The Krafla Power Station was commissioned in 1977, with only unit 1 installed, producing 7 MWe from 11 production wells. Gradually, production was increased and the full capacity of 30 MW was reached when the volcanic eruptions stopped in 1984 and Landsvirkjun (the National Power Company) took over the operation of both power station and the steam supply system in 1985. The following years the reservoir recovered to its pre-eruptional state and the gas content declined. In 1996, the decision was made to install

the second turbine that had been waiting on the shelf for 20 years and the power station was finally on full 60 MW capacity in 1999 (Juliusson et al. 2005).

4.3. Seismicity in Krafla

As Krafla is located above a geothermal area, it is common to observe seismicity which could potentially be related with the geothermal power plant activities. Water extraction and re-injection induces volume changes in the ground, which changes stresses and thus seismic activity associated with it. Natural activity within geothermal areas can also occur, and it is very challenging to distinguish it from the induced activity. Some studies on the Krafla area (Blanck et al. 2016) show that there is a direct relationship between the injection rate and the number of earthquakes recorded (table 4.1). Monitoring this induced seismicity becomes a very important routine to understand how the reservoir is responding to exploitation, and it helps to make it safer and more efficient.

Table 4.1 - Recorded earthquakes related to injection from November 2014 to October 2015 (Blanck et al. 2016).

	Average number of earthquakes per day	**Number of days with 5 earthquakes or more**	**Number of days with 10 or more earthquakes**
Low injection rate (98 days)	2.6	21 (21.65%)	2 (2.06%)
High injection rate (268 days)	4.5	115 (42.91%)	22 (8.21%)

Apart from the seismicity related to the geothermal area, some earthquakes could be also attributed to magma movement, or tectonic stresses as Krafla is located in an active volcanic and tectonic zone. The magma exerts pressure on the rocks changing stresses and producing an earthquake. Then the magma squirts into the crack and starts building pressure again. Normally those earthquakes tend to be weak, but they can be detected and recorded thanks to sensitive instrumentation. As the Krafla region is situated within an active volcanic zone, part of the seismicity recorded in the area could be attributed to magma movement as well.

In the Krafla area the seismic network is composed of 17 stations run by ISOR on behalf of Landsvirkjun, with four of them (HDH, GAESK, HYD and HVE) currently out of service. The SIL network is run by Icelandic Meterological Office (IMO) was not included in this research.

The stations included in this research were installed progressively between 2006 and 2015. Further information about them can be found in table 4.2.

Table 4.2 - Technical information about the stations (Blanck et al. 2016).

STATION	Sensor	Digitalizer	Begin data
HSPHO	Lennartz LE-3Dlite	Reftek	6-jun-2014
SBS	OYO Geospace	Reftek	30-sep-2006
SPB	Lennartz LE-3D5s	Reftek	27-sep-2006
GFJ	Lennartz LE-3D5s	Reftek	30-aug-2013
HVA	Lennartz LE-3D5s	Reftek	30-aug-2013
HVET5	Lennartz LE-3Dlite	Reftek	21-nov-2015
LHN	OYO Geospace	Reftek	14-may-2008
GRT	Lennartz LE-3Dlite	Reftek	29-sep-2006
THEIG	Lennartz LE-3D5s	Reftek	16-oct-2014
BEINI	Lennartz LE-3Dlite	Reftek	16-may-2014
DALFJ	Lennartz LE-3Dlite	Reftek	12-jun-2014
HHK	Lennartz LE-3D5s	Reftek	27-sep-2006
THORF	Lennartz LE-3Dlite	Reftek	1-sep-2014

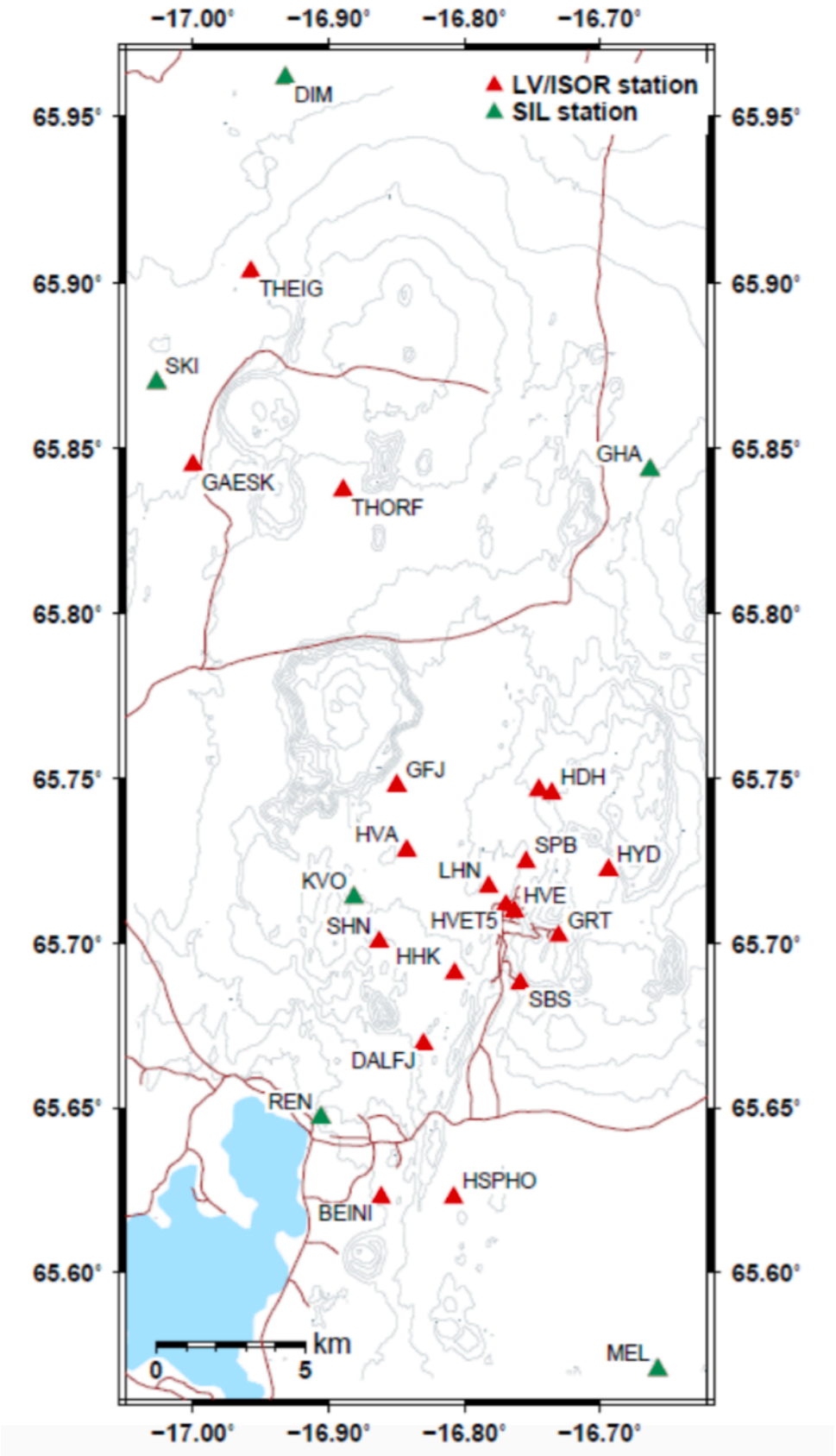

Figure 4.4 - Map of the stations within Krafla and þeistareykir areas (Blanck et al. 2016).

5. SeisComP

SeisComP3 (SC3) is an open source earthquake monitoring system currently used at ISOR with professional software support provided by Gempa GmbH. It follows a modular approach in which every module is a stand-alone program that communicates with other modules through a Transmission Control Protocol/Internet Protocol (TCP/IP) messaging system called Spread and connects to a common database containing the station metadata and other information.

SC3 was designed as a complete Earthquake Early Warning (EEW) solution, capable of estimating early warning magnitudes and epicentral locations. EEW is the concept of using the early observations of seismic energy to predict expected ground-motion levels and provide warning before strong shaking arrives at the same or further-distant sites, often by first estimating event magnitude and location. In practice, no prior component has been implemented, and the location information proved difficult to implement without significant numbers of false alarms when only few stations were used (Cua et al. 2009). For this reason, current implementations of SC3 only provide magnitude estimates and are triggered following independent external fast-event detection and location estimates.

Some of the most relevant SC3 modules for automatic detection and location are described below, as well as in the SC3 processing flow chart (Figure 5.1).

Scenvelope: Produces real time envelope values for horizontal and vertical acceleration, velocity and displacement from raw acceleration and velocity waveforms. It is implemented for the waveform pre-processing necessary for the scvsmag module (see below).

The resulting envelope values are sent as messages to scmaster (see below). The amount of messages depends on the number of streams that are processed, which can result in a considerable number (SeisComP3 documentation).

Scautopick: Searches for waveform anomalies in form of rapid changes in amplitude. It works by applying an STA/LTA algorithm to the waveform streams. Before the algorithm is applied, the waveform is filtered using a Butterworth bandpass filter. The scautopick module is the most important module for the purpose of this research, as the parameters of both the filter and specially the STA/LTA algorithm, can be adjusted. Scautopick allows the possibility of selecting profile configurations which allows to use the same set of parameters for individual stations or group of stations as desired.

Scautopick can run in 2 different modes, online or offline mode. The online mode works in real time, as the module receives stream data. The offline mode processes an input file, using the user-set configuration. This process is called playback and is very useful to test different sets of parameters for the same data set, which is the basic approach applied in this Thesis (SeisComP3 documentation).

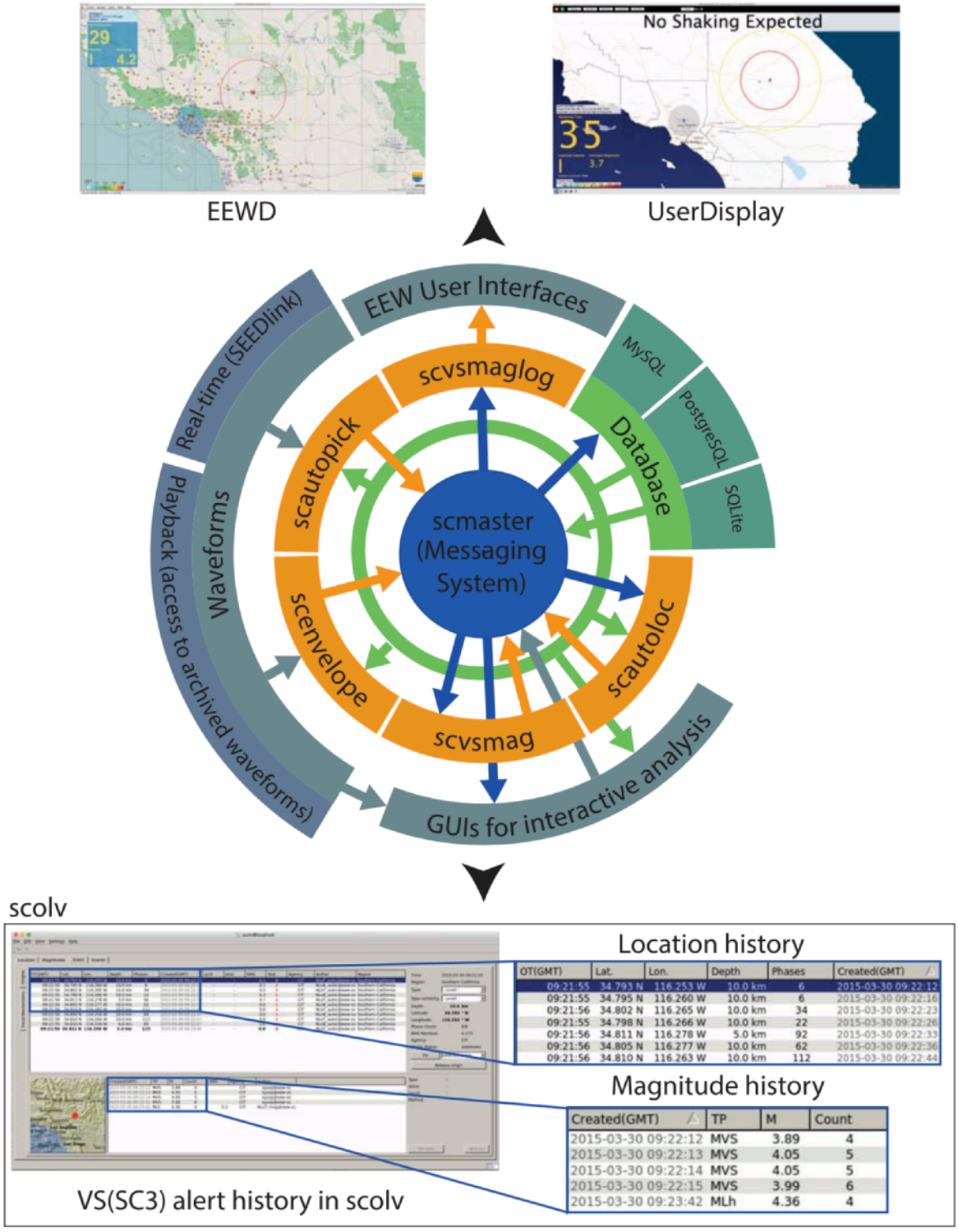

Figure 5.1 - Processing flow chart of SeisComP3. Screenshots are taken from SC3 Graphical User Interface (GUI), scolv (Behr et al. 2016).

Scautoloc: Responsible for automatically locating seismic events. It normally runs continuously reading picks and amplitudes and processing them in real time. Based on that, it tries to identify and relate combinations of picks that correspond to a common seismic

event. The produced location is reported if it meets certain consistency criteria (SeisComP3 documentation).

Scvsmag: Computes magnitudes based on the envelopes calculated by scenvelope. For a given origin, it estimates single station magnitudes and a network magnitude based on the envelope attenuation relationship and ground motion amplitude ratio derived by Cua (2005).

Scmaster: Is responsible of the communication of every single module in SeisComP3. Scmaster was designed as a kind of microkernel which delegates client requests. Thus, it is the key application for the orchestration of the system. In order to participate in the system a client needs to send a connect request to scmaster.

SC3 receives realtime waveform data from seedlink. The first processing modules are scautopick and scenvelope which produce P picks and associated envelopes, respectively. This information is shared to other modules by scmaster. Calculations about event/location and magnitude are done by scautoloc and scmag, respectively. It is possible to visualize and interact with the results thanks to the Graphical User Interface (GUI).

It is important to notice that SC3 has a limitation as it requires at least 6 P phases to create an event. This prevents SC3 from automatically detecting events that are clearly seen in only 3 to 5 stations (SeisComP3 documentation).

6. Methodology

The main objective for this thesis is the selection of optimal STA/LTA parameters in the Krafla area. A brute-force search was applied for all possible STA/LTA parameters within a limited range. Picks for each of the seismic stations are manually selected, and then compared to the output results from the python script described below. Parameters detecting the maximum number of our manual picks and a minimum number of false picks (picks that do not match with the manually selected ones) are considered to be best. This approach offers both simplicity and reliability of the results.

The search was carried out using the open source seismological Python framework "ObsPy" (Beyreuther et al. 2010). This module has a wide variety of useful tools for seismological analysis, including STA/LTA picking tools that also work for SeisComP3. A python script was developed using this tools to iterate over all possible parameter combinations to produce the output results.

6.1. Python script

6.1.1. Selection of the parameters range

The python script is looping over 4 different parameters: short time average (STA), long time average (LTA), the trigger threshold and the detrigger threshold. These parameters are described below.

Taking all the considerations about STA/LTA explained in chapter 3, Trnkoczy (2012) suggests typical values of STA between 1 and 2 seconds when targeting regional events and values between 0.3 and 0.5 for local earthquakes. As the goal of the ISOR network in Krafla is local monitoring in geothermal areas, the smaller values fit better for this purpose. The selected looping values for the STA window length are from 0.1 s to 1 s, with 0.1 s increase (0.1, 0.2, 0.3 … 0.9, 1).

For the purpose of this research short LTA windows are suitable (see chapter 3). The selected values tested for the LTA window are 3 s to 30 s, with an increase increment of 3 s(3, 6, 9, 12 … 27, 30).

The selected values tested for the trigger threshold are 3 to 8 with an increase of 0.5 (3, 3.5, 4 … 7.5, 8). As it is applied to the STA / LTA relationship, the trigger threshold is unitless

The selected values tested for the detrigger threshold are 0.5 to 2.5 with an increase of 0.2 (0.5, 0.7, 0.9 … 2.3, 2.5). As it is applied to the STA / LTA relationship, the detrigger threshold is unitless

6.1.2. Selection of the signal filter

Before going into the looping operation, the trace was filtered using a 4th order Butterworth Bandpass filter, with a minimum frequency of 20 Hz and a maximum

frequency of 50 Hz. The purpose is to filter out frequencies that mostly contain noise and thus to facilitate and improve the accuracy of the triggering routine.

6.1.3. Reference picks and error

For each of the stations, a limited time frame of 2 hours (28-1-2018 15:00 - 17:00) was selected to run the script. First, the waveform for all the active stations within the Krafla network were manually checked and a list of picks obtained. Picks were written as seconds, starting from 28-1-2018 at 15:00. For example, if we have 3 earthquakes at 15:30, 16:00 and 16:30, the input list would be (1800, 3600, 5400).

The script works by comparing the picks obtained manually with the picks obtained automatically from all possible combinations of parameters. However, it is necessary to add an error frame to our manual picks to make sure they are properly identified. The error frame was set to ±1 second. A relatively high value was selected because both errors at manually writing the picks and the automatic detection were considered while still keeping it low enough to avoid associating manual picks with automatic picks which are clearly not from the same event.

The function that calculates the trigger time in Obspy does not use seconds as the default value. Instead, it considers the trace as a succession of data points. These data points correspond to every single one of the sampled values of the trace. For example, for a one second trace at 200 Hz sampling frequency, we would have 200 different points. If this function detects a pick at 1 second, it would record it at “point number 200”. To be able to work with this, we need to multiply all our manual values and error bands by the sampling rate. This value is contained along with the trace, and the conversion can be automatically done with the python script.

6.1.4. Final output

After setting STA and LTA windows, trigger and detrigger thresholds, signal filter, reference picks and error margin, the python script is ready to run. It loops over all possible combinations of STA and LTA windows, trigger and detrigger thresholds for the vertical component of each station and returns a CSV text file with a list of the following elements:

STA: Iterating value

LTA: Iterating Value

TRIG: Iterating value

DETRIG: Iterating value

FALSE: The number of picks that are not within the reference pick plus/minus the error interval

N°picks: Total amount of recorded picks

%: Percentage of correct picks recorded

A reduced sample of random sets of the obtained results is shown in table 6.1.

Table 6.1 - Reduced sample of the obtained results for station HHK.

Input				Output		
STA	**LTA**	**TRIG**	**DETRIG**	**FALSE**	**N°PICKS**	**%**
0.2	15	5	0.5	0	25	100
0.3	9	3.5	1.5	5	30	100
0.6	30	5	0.9	0	21	84
0.9	12	7	0.5	0	10	40
0.5	3	6	0.7	0	0	0
0.1	30	3	2.5	1251	1276	100
0.1	18	5.5	1.1	12	37	100
0.1	6	6.5	1.9	5	29	96
0.2	12	3.5	0.5	17	42	100

6.1.5. Selection of optimal parameters

It is a frequent observation that for every particular station, there are a relatively large number of parameters giving decent results. In SeisComP3, it is possible to add “profile” configuration into the *scautopick* module and this configuration can be applied to multiple stations (see chapter 5). Consequently, it is very convenient to look for sets of optimal parameters that frequently repeat at different stations, or in case that that is not possible, to select a set of parameters which are as similar as possible to the rest of the stations. This would allow to minimize the amount of profile configurations needed, keeping it simple and easy to change in the future, without sacrificing picking efficiency.

First, we delete the worst parameter combinations by ordering the lists obtained from the python script by decreasing percentage of correct picks and increasing number false picks. The set of parameters that do not detect 100% of the correct picks defined before and have a number of false picks of over 30% of the number of real picks were deleted.

The search of similar parameters in different stations was carried out using a simple python script (Appendix 2) which compares the lists of most optimal parameters of two stations and returns another list with parameters in common. This was carried out over all possible

combinations of stations, identifying similar sets of parameters giving sufficient results at different stations.

6.1.6. Comparing the results of ISOR default and optimized detection parameters

The final step is to compare the events detected using both the obtained set of optimized parameters and set of parameters currently used at ISOR over the same time frame. SeisComP3 has the option of running offline playbacks. The playback runs over a given waveform file, using the current SeisComP3 configuration and creates output files containing all picks, magnitudes and events as a result. These files can be loaded into the SeisComP3 database, allowing to visualize in the user interface a list of events, their waveforms on different stations, picks, location and magnitude.

Table 6.2 - Current set of parameters used at ISOR.

STA	LTA	TRIG	DETRIG
0.1	10	3	1.5

After the results are loaded into the database, all the detected events are manually checked and classified between “Real events” and “False events”. An event was classified as “Real event” if it can be observed in at least 4 stations. Some exceptions were applied to this rule, and 3 stations were considered enough for particularly clear and obvious events. An event was classified as “False event” if the previously stated criteria is not met.

7. Results

The results (that is, the selected set of optimal parameters) are presented in table 7.1. We can see that the final results are very similar for all the stations. For each station there was a large number of parameters giving good results and repeating ones were preferred for simplicity reasons, as stated in the previous chapter.

Table 7.1 - Final set of parameters obtained for all active stations within the Krafla area, after following the path described in the previous chapter.

STATIONS	STA	LTA	TRIG	DETRIG
HSPHO	0.2	15	3	0.5
SBS	0.2	15	3	0.5
SPB	0.2	15	3	0.5
GFJ	0.2	15	4	0.5
HVA	0.2	15	4	0.5
HVET5	0.2	15	4	0.5
LHN	0.2	15	4	0.5
GRT	0.2	15	4	0.5
THEIG	0.2	15	4	0.5
BEINI	0.2	15	5	0.5
DALFJ	0.2	15	5	0.5
HHK	0.2	15	5	0.5
THORF	0.2	15	5	0.5

The STA and LTA windows, as well as the detrigger threshold (DETRIG), are the same for all stations and the trigger threshold (TRIG) oscillates between 3 and 5, giving a total of 3 different scautopick profile configurations.

The first profile configuration is characterized by STA window of 0.2, LTA window of 15, trigger threshold of 3 and detrigger threshold of 0.5. It includes the stations HSPHO, SBS and SPB

The second profile configuration is characterized by STA window of 0.2, LTA window of 15, trigger threshold of 4 and detrigger threshold of 0.5. It includes the stations GFJ, HVA, HVET5, LHN, GRT and THEIG

The third profile configuration is characterized by STA window of 0.2, LTA window of 15, trigger threshold of 5 and detrigger threshold of 0.5. It includes the stations BEINI, DALFJ, HHK and THORF.

7.1. Quality check

The quality of the selected parameters can be manually checked by visualizing the waveform along with the produced picks. The quality check was carried out using random time frames, different than the one selected for the parameter optimization to make this quality check more reliable.

In figures 7.1 and 7.3 we can see sections from HVA and HHK stations where a few events are visible, using the values from table 7.1. The STA / LTA ratio shows very clear peaks coinciding with the events, and very stable values between events. This is an ideal situation, because it allows to detect the maximum number of real events and a minimum number of false events. When we take a closer look to a typical event we note that the trigger on and trigger off lines outline the event clearly (Figures 7.2 and 7.4)

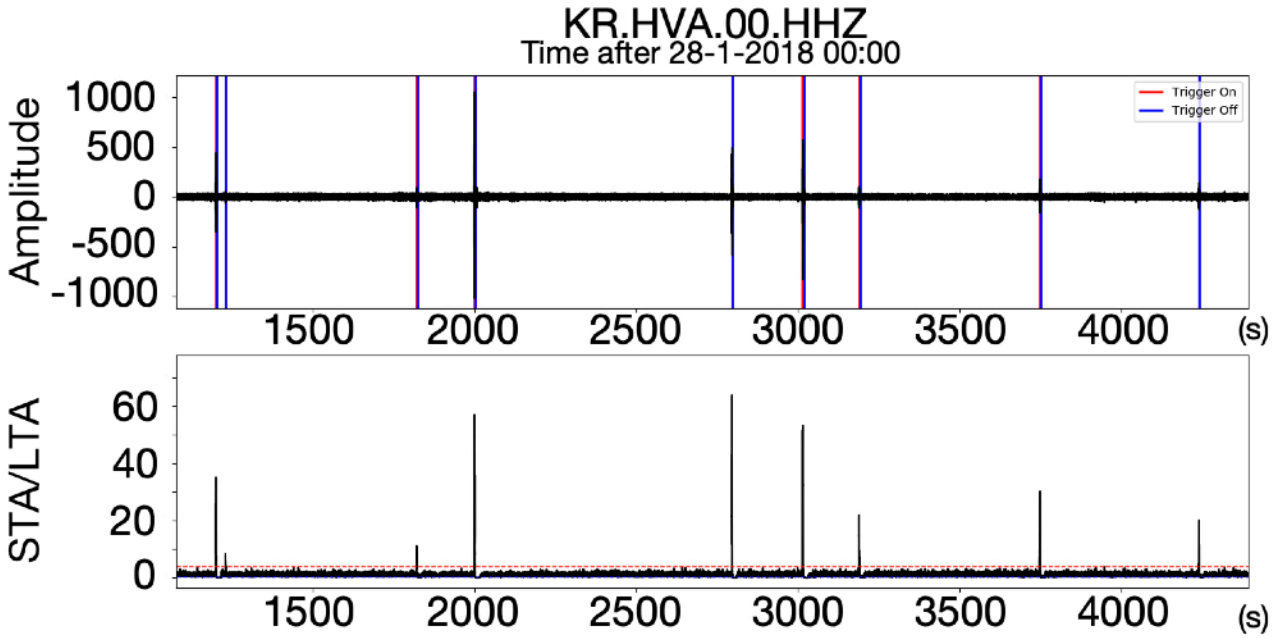

Figure 7.1 - Record section from the HVA station. In the upper figure, the filtered waveform is shown. The vertical red lines represent trigger on (not clearly seen due to overlapping) and vertical blue lines represent trigger off. The lower figure shows the STA/ LTA ratio (horizontal lines represent trigger threshold, and horizontal blue lines detrigger threshold).

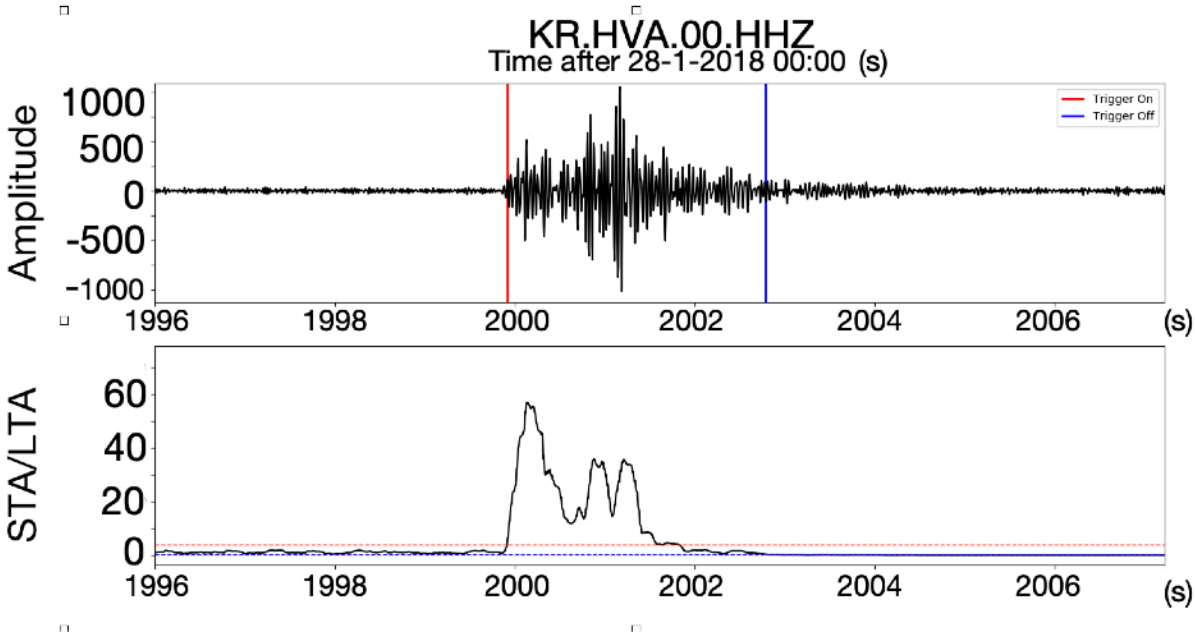

Figure 7.2 - Record section of an earthquake recorded at HVA station. In the upper figure, the filtered waveform is shown (vertical red lines represent trigger on and vertical blue lines trigger off). The figure below shows the STA/LTA ratio (horizontal lines represent trigger threshold and horizontal blue lines detrigger threshold).

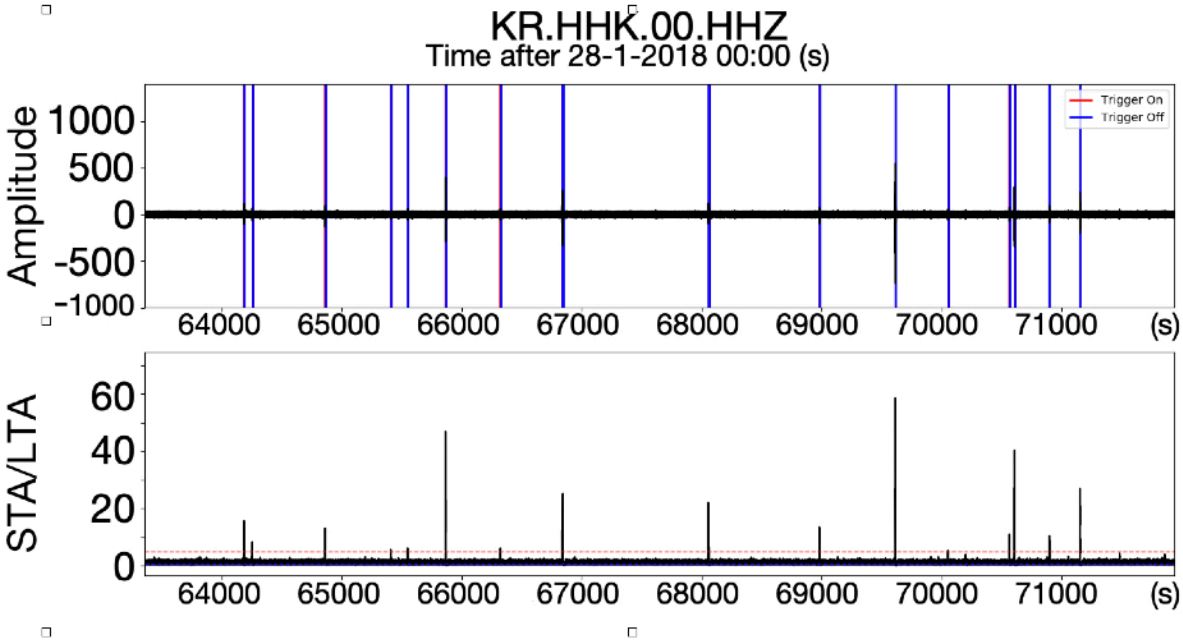

Figure 7.3 - Record section from the HHK station. In the upper figure, the filtered waveform is shown. The vertical red lines represent trigger on (not clearly seen due to overlapping) and vertical blue lines represent trigger off. The figure below shows the STA/LTA ratio (horizontal lines represent trigger threshold and horizontal blue lines detrigger threshold).

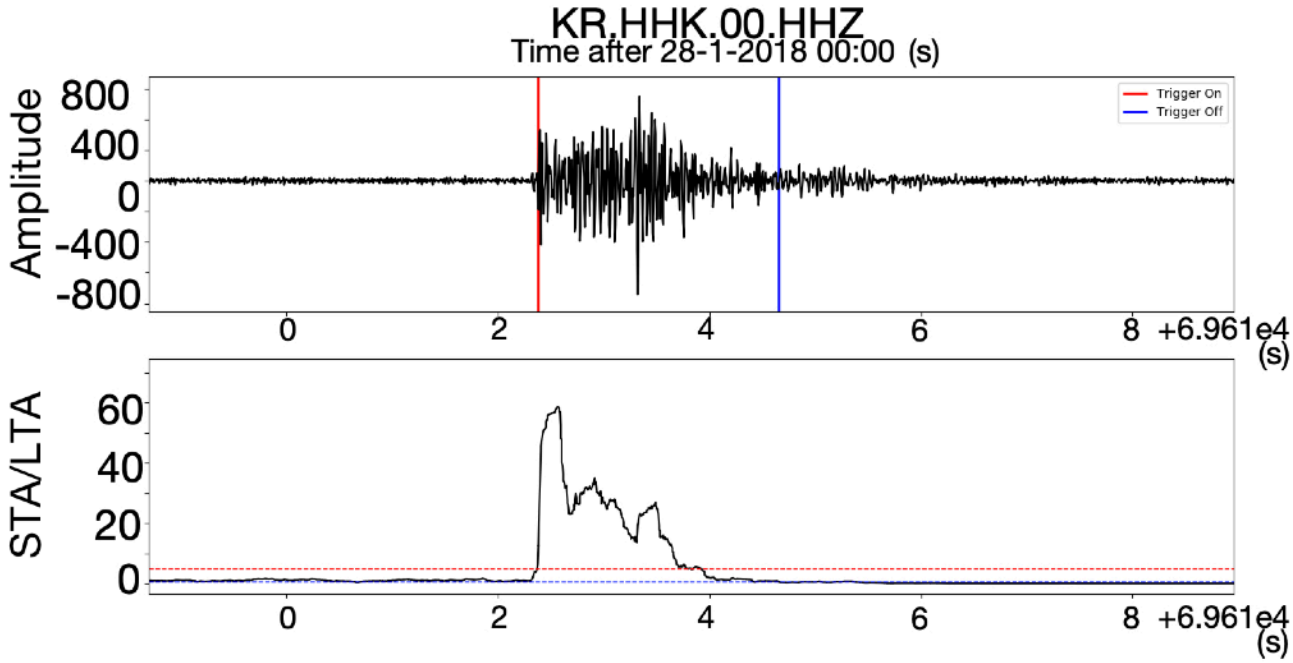

Figure 7.4 - Record section of an earthquake recorded at HHK station. In the upper figure, the filtered waveform is shown (vertical red lines represent trigger on, and vertical blue lines trigger off). The figure below shows the STA/LTA ratio (horizontal lines represent trigger threshold and horizontal blue lines detrigger threshold).

During the signal processing, we apply a Butterworth bandpass filter to increase the signal to noise ratio. Even though this is an effective operation, noise frequency can change drastically with time, making it impossible to get a completely noise-free trace after processing. This can also be observed in our results (see figures 7.5 and 7.6) and it is potentially problematic because it can camouflage a signal from a real event or inducing false triggering as well.

Figure 7.5 shows how the filtered waveform experiences drastic change over a short period (at around 30000 seconds), most likely due to a variation in noise amplitude. This change is also reflected in the STA/LTA radio which should affect the chance of event detection. Figure 7.6 shows a closer look from the noisy area in figure 7.5. The trace was checked in detail and only two small events were manually identified at around 30900 seconds that were not automatically detected.

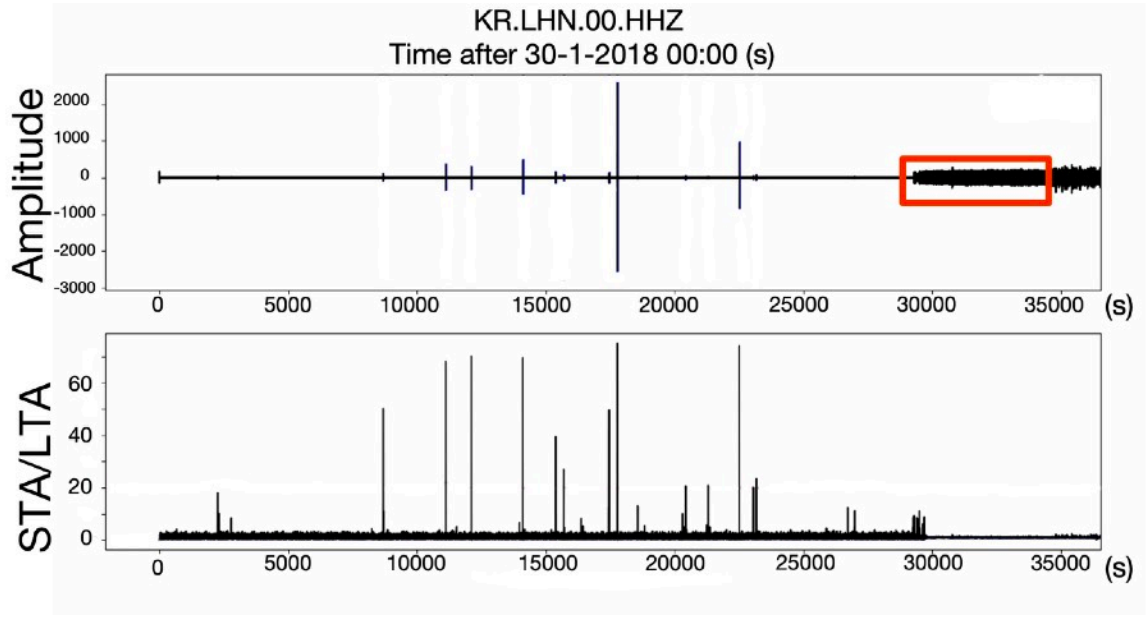

Figure 7.5 - Record section from the LHN station. In the upper figure the filtered waveform is shown. The figure below shows the STA/LTA ratio. The red box shows a noisy area. examined in figure 7.6

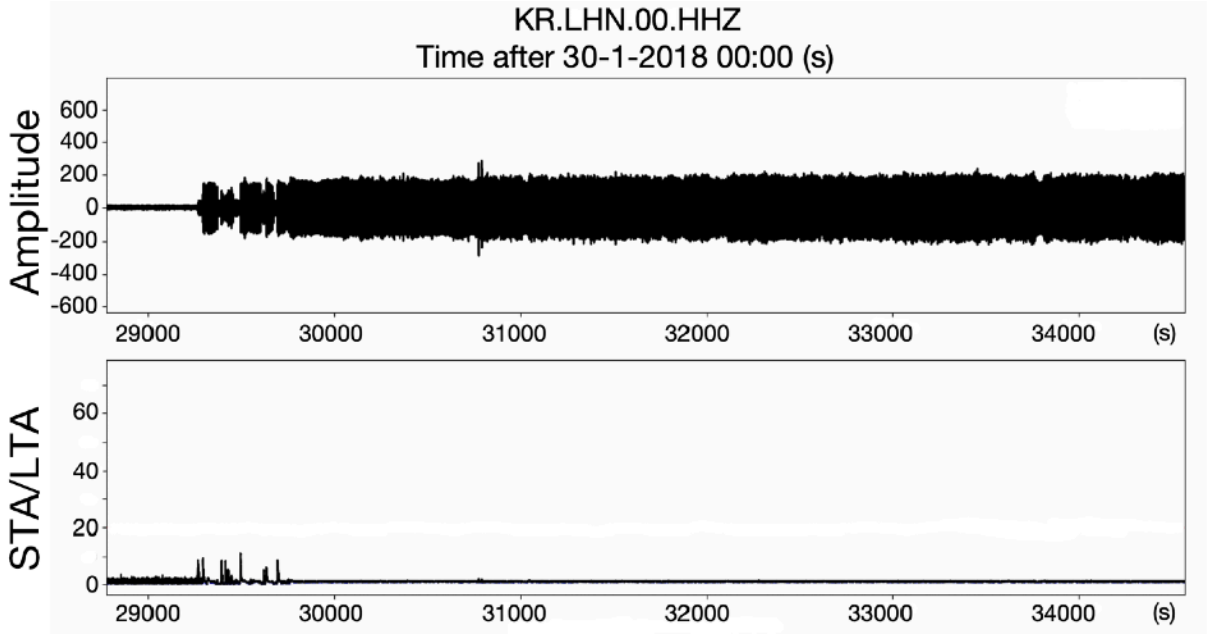

Figure 7.6 - Closer look over the noisy area from HHK station. In the upper figure the filtered waveform is shown. The figure below shows the STA/LTA ratio.

Table 7.2 shows a comparison of the number of real and false events using the new and old sets of parameters in SC3. We can see that the new parameters are not recording any false event, but they are only detecting half of the real events recorded using the old parameters. It is interesting to point out that all the events recorded by the new parameters were also recorded by the old parameters. These results are surprising at first sight, but they can be

explained by the SeisComp limitation of minimum P phases (see SeisComP3 and Discussion chapters for further details).

Table 7.2 - Comparison between results obtained using the new and the ISOR default parameters.

	Real events	**False events**	**Total**
New parameters	28	0	28
ISOR default	62	16	78

8. Discussion

When we look at the results from table 7.2, there are 2 factors that have to be discussed:

1. Even though there are no false events recorded using the new parameters, the amount of real events is considerably lower than using the old parameters. This could be explained by the SeisComP3 limitation of minimum phases. SeisComP3 needs at least six P-phases to recognize an event. This means, that the event has to trigger six or more stations, otherwise it won't be detected. However, six stations are still too high, as we can properly identify a real event if it is observed in only three to five stations. This is a frequent observation in the events recorded by the old parameters and omitted by the new ones. In figure 8.1, we can see an example from an event detected with the old parameters that it is only visible in 5 stations. The event was detected because false picking completed the number of minimum P-phases. As the new parameters are producing less false picking they are unable to reach the limit and the event is missed. As a consequence, the total number of real events detected using the new parameters is considerably lower.

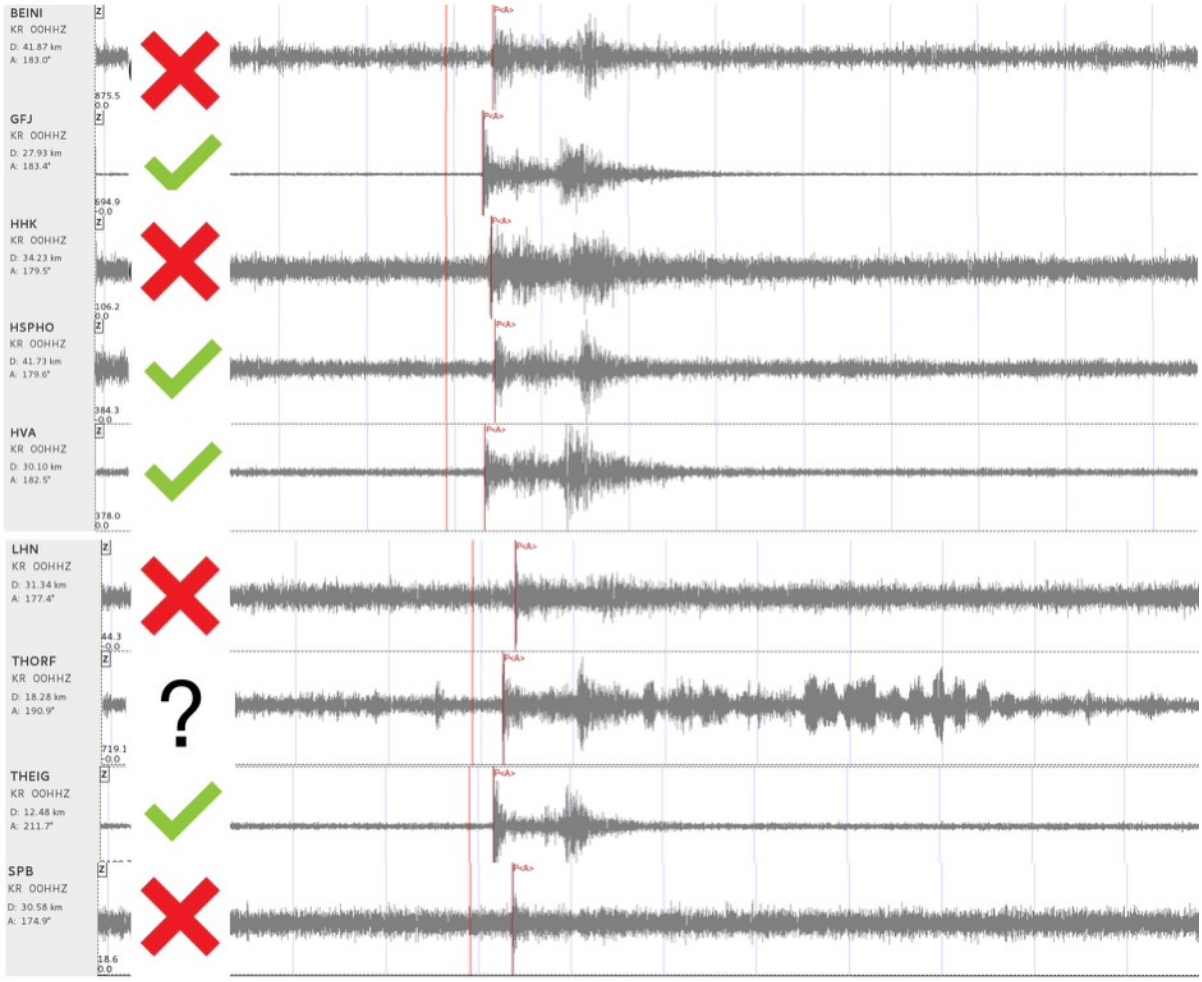

Figure 8.1 - Example of an event that was omitted using the new parameters and detected by the old ones. The green ticks represent the traces where the event is clear and the red crosses represent those traces where it is not, and could be omitted by SeisComP3.

It is important to point out that this SeisComP3 limitation makes it impossible to simulate what would happen if the limitation itself disappeared. It would be logical to think that most of the events omitted by the new parameters and detected by the old ones (like the example in figure 8.1) could be detected, and many false events would be omitted as the picker (Scautopick) is producing less false picking. However, the impossibility of testing it directly makes this conclusion hypothetical.

2. When we look at the output values from the python script (see table 6.1) a trend can be observed regarding the number of picks and the percentage of good picks. As the number of recorded picks increase, so does the percentage of correct ones. It will eventually reach an optimal status, after which only the number of false pickings will increase. This trend is represented in figure 8.2.

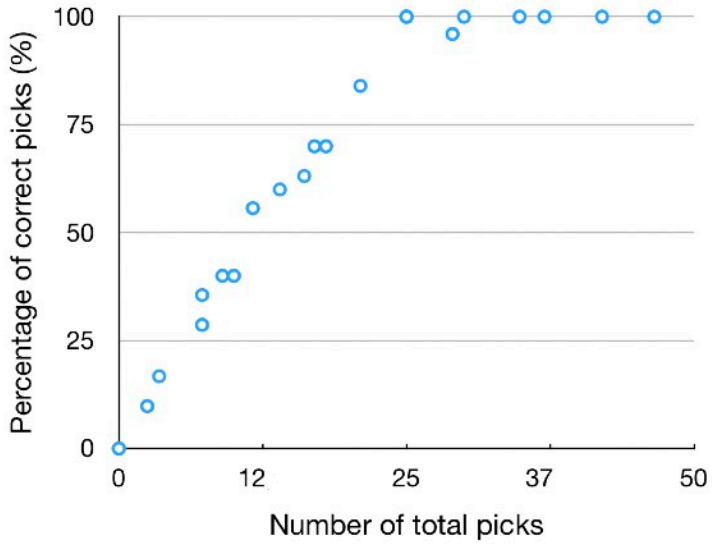

Figure 8.2 - Relationship between the percentage of correct picks and the number of total picks. Each point represents a set of parameters, although different set of parameters can sometimes produce the same results. This data comes from HVA station, where the number of manually selected correct picks was 25.

This is suggesting a framework (represented in figure 8.3), in which every single set of parameters can be projected and catalogued into 3 categories:

- Under-picking: Includes all sets of parameters with low number of recorded picks. They are characterized by the absence of most events (both real and false). They may have some practical application for targeting only the strongest events, but it would be unpractical for a local seismic monitoring, as most of the real events will be missing. Some examples of this type of parameters are presented in table 8.1

Table 8.1 - Example of underpicking set of parameters from station HHK.

STA	LTA	TRIG	DETRIG	FALSE	N°PICKS	%
0.3	3	4.5	0.5	0	17	70
1	27	6.5	1.7	0	14	60
0.3	3	7	0.5	0	9	40
1	3	4.5	2.1	0	0	0
0.9	30	5.5	1.1	1	18	70

- Optimal picking: Includes all sets of parameters with a higher real/false picking ratio. Under optimal conditions (without the SeisComP3 limitation of minimum phases) they would be the most practical for targeting all kind of events as they detect most of the real events while offering minimum false triggering. At the same time, a more accurate and reliable picking would also increase the accuracy of location and depth calculations.

Table 8.2 - Example of optimal picking set of parameters from station HHK.

STA	LTA	TRIG	DETRIG	FALSE	N°PICKS	%
0.2	15	5	0.5	0	25	100
0.2	60	6	0.9	0	25	100
0.4	9	3.5	1.1	1	25	100
0.1	18	7.5	0.9	2	25	100
0.3	9	4	1.1	1	25	100

- Over-picking: Includes all sets of parameters with a lower real/false picking ratio. They could be useful to make sure no real event is missed. As a drawback they produce false triggering, and human review is necessary to discriminate between real and false events. The accuracy of calculations for location and depth would also be affected, as false picks could be considered as real (like in figure 8.1).

Table 8.3 - Example of an overpicking picking set of parameters from station HHK.

STA	LTA	TRIG	DETRIG	FALSE	N°PICKS	%
0.1	30	3	1.3	1172	1197	100
0.2	30	3	2.5	98	123	100
0.1	18	4	2.5	126	151	100
0.1	30	3.5	1.7	334	359	100
0.1	3	3	0.7	845	870	70

The new set of parameters obtained could be catalogued under the "optimal picking" category, while the current parameters used at ISOR fit better under "over-picking". It is important to notice that this is a theoretical classification. Theoretically optimal parameters may not be optimal in practice. The goal at ISOR is to detect as many events as possible, so the results are suggesting that the new set of parameters are not a practical choice, due to the SeisComP3 limitation.

Considering this scenario, there are 2 possible solutions:

- The first solution is to modify the SeisComP3 code and eliminate the restriction of minimum P phases. As explained in the example of figure 8.1, it would allow to preserve most of the real events and considerably reduce the number of false ones.

- If the first option is not available, then the current set up, with "over-picking" parameters is a good option. We would record most of the real events, but at the cost of accepting a larger number of false ones, and manual re-picking would be necessary for accurate magnitude and depth estimations as well.

However, inside the over-picking category there is some variability. A set of parameters that is located in the right extreme of the scale (figure 8.4) would detect more false events that one located closer to the center. To test where in the scale the current parameters used at ISOR are located, one more test was done, slightly changing the triggering value (+-1) and analyzing the results. The current set of parameters used at ISOR is the same for all stations within the Krafla network (table 8.4)

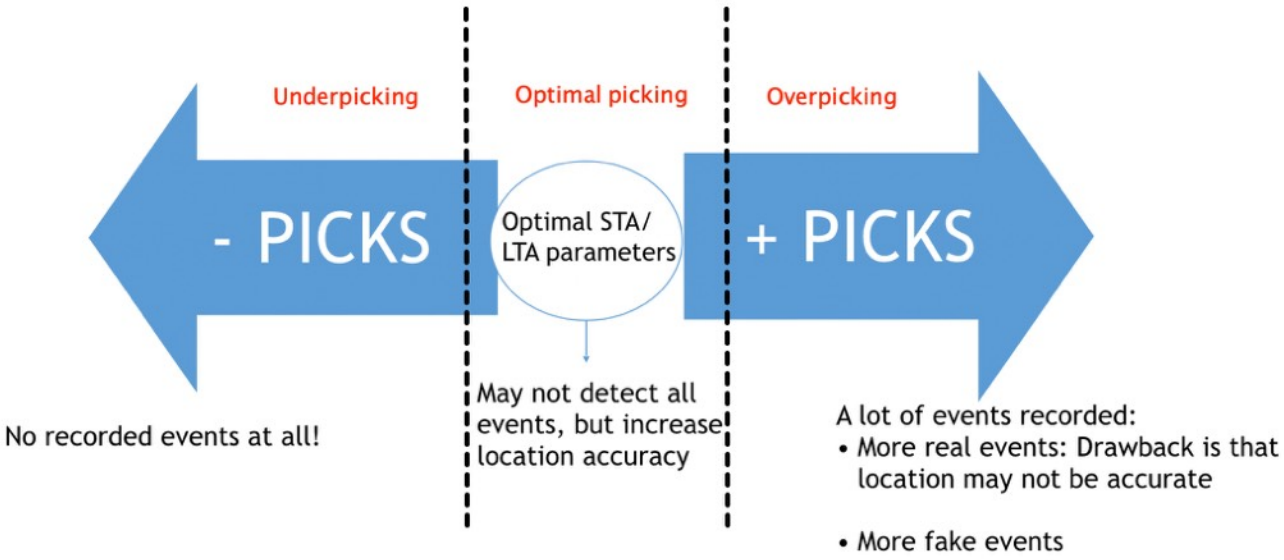

Figure 8.4- Visualization of the picks framework.

The results obtained after increasing and decreasing the trigger value from the current set of parameters at ISOR is presented in table 8.5.

Table 8.5 - Number of events recorded after decreasing and increasing the trigger value.

	Real events	**False events**	**Total**
Trigger value lowered to 2	63	28	91
Trigger value increased to 4	50	12	62

We can see that when we decrease the trigger value the number of real events is about the same as ISOR default parameters (62 compared to 63), while the number of false events increases around 75% (16 compared to 28). When the trigger value is increased, the number of real events decreases around 19% (62 compared to 50), and the number of false events decreases around 25% (16 compared to 12).

The results presented above are suggesting, that under the current circumstances and goals, the set of parameters used by ISOR is appropriate. However, significant improvements could be done if SeisComP3 limitation could be bypassed.

9. Conclusions

The main goal of this research was to improve the automatic earthquake detection system of ISOR at the Krafla geothermal area in North Iceland. It was carried out by modifying the variables from the STA/LTA algorithm used by SC3.

We find that it is possible to find a theoretically enhanced set of STA/LTA variables for the Krafla network. But at the same time SC3 has a limitation (minimum number of P phases) that makes this set of parameters difficult to implement. However, there are definitely some ideas to improve the current detection system. For example, removing the SC3 limitation mentioned would allow significant improvements, specially reducing the number of false events.

In the case of the ISOR seismic network, similar experiments on improving the earthquake detection systems could be tried in the rest of the seismic networks as well, although there is a chance that the same handicap will appear unless SC3 restriction is bypassed.

Running the experiment over a longer time window is an interesting alternative as weather factors would have a lower impact. However, it also increases the amount of manual picking required, making it harder to handle.

References

Behr, Y., J. F. Clinton, C. Cauzzi, et al. (2016). "The Virtual Seismologist in SeisComP3: A New Implementation Strategy for Earthquake Early Warning Algorithms." Seismological Research Letters **87**.

Beyreuther, M., R. Barsch, L. Krischer, et al. (2010). "ObsPy: A Python Toolbox for Seismology." Seismological Research Letters **81**(3): 530-533.

Bjarnason, I. þ. (2008). "An Iceland hotspot saga."

Blanck, H., K. Ágústsson and K. Gunnarsson (2016). Seismic Monitoring in Krafla. November 2014 to November 2015.

Brown, J. R., G. C. Beroza and D. R. Shelly (2008). "An autocorrelation method to detect low frequency earthquakes within tremor." Geophysical Research Letters **35**(16).

Cua, G., M. Fischer, T. Heaton, et al. (2009). "Real-time Performance of the Virtual Seismologist Earthquake Early Warning Algorithm in Southern California." Seismological Research Letters **80**(5): 740-747.

Demets, C., R. G. Gordon, D. F. Argus, et al. (1994). "Effect of Recent Revisions to the Geomagnetic Reversal Time-Scale on Estimates of Current Plate Motions." Geophysical Research Letters **21**(20): 2191-2194.

GmbH, g. SeisComP3 documentation.

Hjartardottir, A. R., P. Einarsson, E. Bramham, et al. (2012). "The Krafla fissure swarm, Iceland, and its formation by rifting events." Bulletin of Volcanology **74**(9): 2139-2153.

ISOR. (2019). "ISOR website." from www.isor.is.

Juliusson, B. M., B. Palsson and A. Gunnarsson (2005). "Krafla Power Plant in Iceland – 27 years of operation." Proceedings of the World Geothermal Congress.

Loftsdottir, A. and R. Thorarinsdottir (2006). Energy of Iceland. Historical perspective, present status, future outlook, Orkustofnun.

Pope, E. C., D. K. Bird, S. Arnorsson, et al. (2016). "Hydrogeology of the Krafla geothermal system, northeast Iceland." Geofluids **16**(1): 175-197.

Schaff, D. P. and G. C. Beroza (2004). "Coseismic and postseismic velocity changes measured by repeating earthquakes." Journal of Geophysical Research-Solid Earth **109**(B10).

Schuler, J., T. Greenfield, R. S. White, et al. (2015). "Seismic imaging of the shallow crust beneath the Krafla central volcano, NE Iceland." Journal of Geophysical Research-Solid Earth **120**(10): 7156-7173.

Tarbuck, E. J., F. K. Lutgens and C. J. Tsujita (2005). Earth : an introduction to physical geology. Toronto, Pearson Prentice Hall.

Trnkoczy, A. (2012). Understanding and parameter setting of STA/LTA trigger algorithm. New Manual of Seismological Observatory Practice 2 (NMSOP-2)**:** 20 pp.

Wang, Z. J. and B. M. Zhao (2017). "Automatic event detection and picking of P, S seismic phases for earthquake early warning and application for the 2008 Wenchuan earthquake." Soil Dynamics and Earthquake Engineering **97**: 172-181.

Withers, M., R. Aster, C. Young, et al. (1998). "A comparison of select trigger algorithms for automated global seismic phase and event detection." Bulletin of the Seismological Society of America **88**(1): 95-106.

Yoon, C. E., O. O'Reilly, K. J. Bergen, et al. (2015). "Earthquake detection through computationally efficient similarity search." Science Advances **1**(11).

Appendix 1

In this appendix, the python script used for the search of optimal set of parameters is presented.

```
“""

Created in 2019 (Version 1.0)

@author: Iñigo Sevilla Echeverría

"""

#AUTOMATIC SEARCH FOR OPTIMAL STA/LTA PARAMETERS

# Import all the necessary libraries

from obspy.signal.filter import bandpass

from obspy import read

from obspy.signal.trigger import classic_sta_lta, trigger_onset

import csv

# This is the place where the minimum values are set

STA_MIN =  #Values mentioned in text were set here. Could be adjusted as desired

LTA_MIN =  #Values mentioned in text were set here. Could be adjusted as desired

TRIG_MIN =  #Values mentioned in text were set here. Could be adjusted as desired

DETRIG_MIN =  #Values mentioned in text were set here. Could be adjusted as desired

# This is the place where the maximum values are set

STA_MAX = #Values mentioned in text were set here. Could be adjusted as desired

LTA_MAX = #Values mentioned in text were set here. Could be adjusted as desired

TRIG_MAX = #Values mentioned in text were set here. Could be adjusted as desired

DETRIG_MAX = #Values mentioned in text were set here. Could be adjusted as desired

# N° of iterations per parameter. It allows to have a good control over the number of total iterations, and thus the total time the script will take to run.

NSTA =  #Could be adjusted as desired
```

```
NLTA =  #Could be adjusted as desired

NTRIG =  #Could be adjusted as desired

NDETRIG =  #Could be adjusted as desired

# Waveform data

st = read('trace file')

# Index of the selected channel

n=0

df = st[n].stats.sampling_rate

# Filtering the trace

tr = bandpass(st[n].data, 20, 50, df, corners=4)

st[n].data = tr

#Reference picks and maximum allowed error band

picks = []

ERROR =

#Convert the pick times into sample numbers to avoid bugs with OBSPY

def sample(p):
    for i in range(0,len(p)):
        p[i]=df*p[i]
    return p

picks = sample(picks)

ERROR = ERROR * df

# Setting the initial iterating parameters

STA = STA_MIN

LTA = LTA_MIN

TRIG = TRIG_MIN
```

```
DETRIG = DETRIG_MIN

#List of final parameters. Values will be adding here as they are calculated, and later exported to csv file for the final output

param = []

# Looping of STA/LTA parameters

while STA <= STA_MAX:

    while LTA <= LTA_MAX:

      while TRIG <=TRIG_MAX:

        while DETRIG <= DETRIG_MAX and DETRIG <= TRIG:

#The following bold code is the core of the script. STA/LTA function is calculated and trigger times computed. After that, trigger times are compared with the manual picks and the results written in “param” list

            cft = classic_sta_lta(st[n].data, int(STA * df), int(LTA * df))

            on_off = trigger_onset(cft, TRIG, DETRIG)

            time = []

            for i in on_off:

               time.append(i[0])

            correct_picks = 0

            for pick in picks:

               for element in time:

                  if element >= pick and element <= pick + ERROR:

                     correct_picks += 1

                     break

            FALSE = len(time) - correct_picks

                     p = [STA, LTA, TRIG, DETRIG, FALSE, len(time), correct_picks/ len(picks)]

            param.append(p)
```

```
                DETRIG += (DETRIG_MAX - DETRIG_MIN)/(NDETRIG - 1)
            TRIG += (TRIG_MAX - TRIG_MIN)/(NTRIG - 1)
          DETRIG = DETRIG_MIN


        LTA += (LTA_MAX - LTA_MIN)/(NLTA - 1)
        TRIG = TRIG_MIN
    STA += (STA_MAX - STA_MIN)/(NSTA - 1)
    LTA = LTA_MIN
#Export final results into csv text file
text = 'output file'
with open(text, "w") as output:
    writer = csv.writer(output, lineterminator='\n')
    writer.writerows(param)
```

Appendix 2

In this appendix, the python script used to search for repeating set of parameters at different stations is presented.

```
“"""

Created in 2019 (Version 1.0)

@author: Iñigo Sevilla Echeverría

"""

#List of optimal parameters for both stations go here

station1 = []

station2 = []

#Final list of common parameters between the stations

comparison = []

#Main body of the code. The comparison between both stations is made, and common parameters added into the final list

for i in station1:

    if i in station2:

        comparison.append(i)

#Parameters in common displayed

print(comparison)
```